

TEACHER COLLABORATION AND STUDENT OUTCOMES IN SAUDI
ARABIA: AN ANALYSIS OF TIMSS DATA

By

© 2018

Mohammed Hindi Alghamdi

M.A., King Saudi University, 2000
B.Sc., King Khalid University, 1993

Submitted to the graduate degree program in Educational Leadership and Policy
Studies and the Graduate Faculty of the University of Kansas in partial
fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Michael Imber

Lesa Hoffman

Argun Saatcioglu

Thomas DeLuca

Joe Novak

Date Defended: March 2nd 2018

The Dissertation Committee for Mohammed Alghamdi certifies
that this is the approved version of the following dissertation:

TEACHER COLLABORATION AND STUDENT OUTCOMES IN SAUDI
ARABIA: AN ANALYSIS OF TIMSS DATA

Chair: Michael Imber

Date Approved: March 2nd, 2018

ABSTRACT

This study was conceptually built on the premise that an internal change that is created by the knowledge and expertise of teachers is more likely to nurture a successful educational reform. To empirically examine this premise, the study investigated the direct and indirect relationship between teacher collaboration (CIT) and student outcomes. The indirect relationship was mediated by two teaching practices, teacher Emphasis on Academic Success (ACS) and their Instruction to Engage Student in learning (IGS). The data was obtained from TIMSS-2011 of Saudi 4th grade students, their teachers, and school principals. Using Structural Equation Modeling (SEM), the study found neither direct nor indirect significant relationship between teacher collaboration and student outcomes. Interestingly, teacher collaboration was negatively and significantly associated with both mediators. Several implications for policy makers and practitioners as well as future research were discussed.

Acknowledgments

Throughout this journey, I was fortunate to work with talented people towards greater understanding and appreciation for the profession of education. To Dr. Michael Imber, I am indebted to this talented mentor who has inspired me to deeply examine school reform initiatives and who continued to push me. To Dr. Lesa Hoffman, this study could not have been developed methodologically without your great mentorship, support, and patience. I am in gratitude to Dr. Argun Saatcioglu for his support and informative thoughts through different stages of this project. I also extend gratitude to the other dissertation committee members, Dr. Thomas DeLuca and Dr. Joe Novak, for their time, effort, and feedback.

Dedication

I have been blessed to have tremendous support from my immediate and extended family who has encouraged me during this journey. This dissertation and degree are dedicated to the memory of my father, Hindi, and my brother, Majed. To my mother, Reda, for her unprecedented prayers. To my wife, Salihah, for her outstanding support and patience. To my sons, Anas, Majed, Mazen, Yazan, Qais, and my daughter Yara, for their understanding, support, and love through this project. My dedication is extended to my brother, Ateg, for his commitment and support and to my other brothers and sisters. I would like to extend thanks to my friends who have encouraged me and helped me accomplish this mission.

TABLE OF CONTENTS

ABSTRACT.....	iii
Acknowledgments.....	iv
Dedication.....	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
CHAPTER 1 INTRODUCTION	1
1.1 Overview.....	1
1.2 Study Purpose.....	4
1.3 Conceptual Framework.....	5
1.4 Study Problem.....	5
1.5 Study Hypotheses.....	6
1.6 Contribution to Literature.....	7
1.7 Significance of the Study.....	8
1.8 Study Rational.....	9
1.9 Overview of Dissertation.....	10
CHAPTER 2 THEORITICAL FRAMEWORK	11
2.1 Teacher Collaboration and Adaptive Expertise.....	11
2.2 School Contexts and Teachers' Learning Opportunities.....	14
2.3 Collaborative School Culture.....	20
2.4 Teacher Change.....	24
2.5 Student Engagement.....	27
2.6 Student Confidence in Learning.....	29
2.7 Educational Reform Profile in Saudi Arabia.....	29
CHAPTER 3 METHOD	45
3. 1 Overview.....	45
3. 2 Target Population.....	46
3. 3 Sample.....	46
3. 4 Data Collection.....	47
3. 5 Study Variables and Instruments.....	47
3. 6 Analysis Software.....	54
3. 7 Analysis Methods.....	55
CHAPTER 4 STUDY FINDINGS	60

4. 1. Measurement Model	61
4. 2 Structural Model	66
CHAPTER 5 DISCUSSION	72
5. 3 Implications.....	79
5. 4 Limitations and Future Research	82
REFERENCES	84
Appendixes	93
<i>Appendix 1.</i>	93
<i>Appendix 2.</i>	93
<i>Appendix 3.</i>	93

LIST OF TABLES

Table 1: Strategic Objective 4 of the National Transforming Program (NTP), one of the 2030's Saudi Vision Programs.	37
Table 2: Interclass Correlation for the Highest and Lowest Level One Construct Items..	62
Table 3: Measurement Models.....	63
Table 4: Correlation between Constructs in L1	65
Table 5: Correlations between Constructs in L 2.....	65
Table 6: Mediated Relationships Between CIT and ENG, CON, and PVM	78

LIST OF FIGURES

Figure 1. The Hypothesized Relationship between the Study Constructs.....	5
Figure 2. Two dimensions of learning and transfer: and the optimal adaptability corridor.....	15
Figure 3. The Structural Relationships between the Study Constructs.....	58

CHAPTER 1

INTRODUCTION

1.1 Overview

In many educational systems around the world, reformers have advocated for improving teaching quality that would eventually lead to improved student outcomes (Cuban, 1996; Darling-Hammond & Wei, 2009; Goddard, Goddard, & Tschannen-Moran, 2007; Waldron & McLeskey, 2010). Yet, improving classroom teaching has been a complex task for any reform initiative (Tyack & Cuban, 1995; Tyack & Tobin, 1994). While it is an established sociological tenet that complex tasks require comprehensive efforts to accomplish, most education organization is fortified against collaborative work for both architectural and social reasons (Little, 1999). Architecturally, Lortie (1975) was the first to observe that the “egg crate” structure of schools has played a crucial role in overstating the fundamental principal of teaching autonomy. Consequently, this school structure (autonomy) has led to another social phenomenon in teaching, which was named by Huberman (1995) as a “lone wolf scenario”. In this most common scenario, teachers work on their own to decide what instruction, under what standards, and based on what knowledge and skills would best serve their students (Bartalo, 2009; Darling-Hammond & Wei, 2009; Little, 1982).

The isolated conditions of teachers’ work have been recognized as an important barrier to improving teachers’ professional learning opportunities. This have encouraged reformers to suggest more deliberated support of teachers’ work that would enhance their teaching quality. While Little (1982) for example, asserted that there is neither a simple checklist nor a ready-made template that schools could employ to support teacher learning (in McLaughlin & Marsh,

1990), he admitted that “joint work” is broadly conceived as an opportunity for teacher learning (p. 234). Later, Tyack and Cuban (1995) also reached a similar conclusion when they underscored that close collaboration among teachers that facilitates and supports internal change is not only necessary to any initiative reform to achieve its goals, but it also is considered a broad-based social movement necessary to any institutional reform. Since then, researchers have looked at effective school practices to evaluate their impact on teaching quality and students learning. In their comparative study using TIMSS Video Study, Stigler and Hiebert (2000), for instance, concluded that despite teachers’ efforts in implementing reforms, measures, and recommendations, little evidence has been provided that indicates changes in U. S. schools. Moreover, they cautioned that unless the teachers’ culture of learning is changed, student performance in U.S. schools would continue to lag behind expectations. Additional evidence has come from the Program for International Student Assessment (PISA). After reviewing teacher practices in countries that ranked at the top of this program, an OECD report found that teachers in these countries are expected to contribute to the knowledge base of effective teaching, be involved in a collaborative process of lesson plans, and observe their colleague’s practices during their teaching (Schleicher, 2012).

Recently, more emphasis has been placed on teacher collaboration (Goddard, Goddard, Larsen, & Jacob, 2010; Goddard et al., 2007; Hargreaves & Fullan, 2013; Ronfeldt, Owens, McQueen, & Grissom, 2015; Miller, Waldron & McLeskey, 2010). Morse (2000), for example, emphasized that in any educational reform, collaboration between practitioners is imperative and is not an option. Other researchers asserted that collaborative work environment becomes the norm of successful organizations (Darling-Hammond & Wei, 2009; Vangrieken, Dochy, Raes, & Kyndt, 2015). They underlined that collaboration in education is not only necessary for educators

themselves but also for their students. That is, while proficient collaboration between teachers is required to implement innovative and student-centered learning methods, it plays a critical role in preparing students to be proficient future collaborators (Coke, 2005; Vangrieken et al., 2015).

The growing significance of teacher collaboration in school effectiveness has encouraged other researchers to investigate its impact on school outcomes at different levels. In a pioneer study, Little (1982) concluded that collegiality and work together between teachers were the two most distinguished factors of school success in his school samples. Other studies found that the quality of the relationships among staff members in a school is a signal of effective organizational culture (Kelchtermans, 2006; Peterson & Brietzke, 1994; Sammons, Hillman, & Mortimore, 1995; Vangrieken et al., 2015). At the teacher level, other researchers have concluded that using more innovative pedagogies, better job satisfaction, self-efficacy, and higher levels of trust are among many positive outcomes of teacher collaboration (Goddard et al., 2007; Vangrieken et al., 2015). At the student level, several studies (Goddard et al., 2007; Goddard, Miller, Larson, & Goddard, 2010; Moolenaar, Daly, & Slegers, 2010; Ronfeldt et al., 2015) found preliminary evidence that teacher collaboration enhances student performance, specifically their achievement as the most significant school outcome.

While these studies have reported on the positive impact of formal teacher collaboration, other studies have, however, doubted the possibility of improvement that collaboration might bring about to schools that have been characterized historically by isolated work modes (Goddard et al., 2007; Lortie, 1975; Tyack & Cuban, 1995; Waldron & McLeskey, 2010). Kelchtermans (2006) called for a certain level of sophistication to understand the effects of teacher collaboration on schools. He cautioned against the “simplistic claims about the benefits of collaboration are as little warranted as negative judgments about teacher autonomy” (p. 234).

Kelchterman's conclusion was supported by Goddard et al. (2007) when the latter asserted that while "collaboration is often advocated, yet its effects are less frequently investigated" (p. 878). Other scholars indicated that empirical evidence that links directly teacher collaboration to student outcomes is still limited (Goddard et al., 2007; Nienke, Moolenaar, Slegers, & Daly, 2012; Ronfeldt et al., 2015).

1.2 Study Purpose

Based on the evidence currently available, the direct link between teacher collaboration and student outcomes has not been well established (Goddard et al., 2007; Ronfeldt et al., 2015; Vangrieken et al., 2015). Therefore, this study will consider the limited existing knowledge about teacher collaboration and its role in improving student outcomes. Specifically, this study empirically investigates the relationship between teacher collaboration and student engagement, confidence in learning, and student achievement in mathematics. The data was utilizing from the Trends in International Mathematics and Science Study (TIMSS, 2011) for fourth grade students in Saudi Arabia.

1.3 Conceptual Framework

The theoretical premise behind this study is based on Tyack and Cuban's (1995) conclusion that "to bring about improvement at the heart of education, classroom instruction will not be achieved unless an internal change created by the knowledge and expertise of teachers are facilitated and supported" (p. 135). They elaborated that close collaboration between practitioners who share common purposes would help give them flexibility in adapting reforms to their local circumstances. These collaboration efforts might be regarded as a broad-based social movement necessary to any institutional reform. Therefore, this study is an attempt to empirically examine the extent to which teacher collaboration, as a way of an internal change, would improve student outcomes as an ultimate goal of any school reform effort.

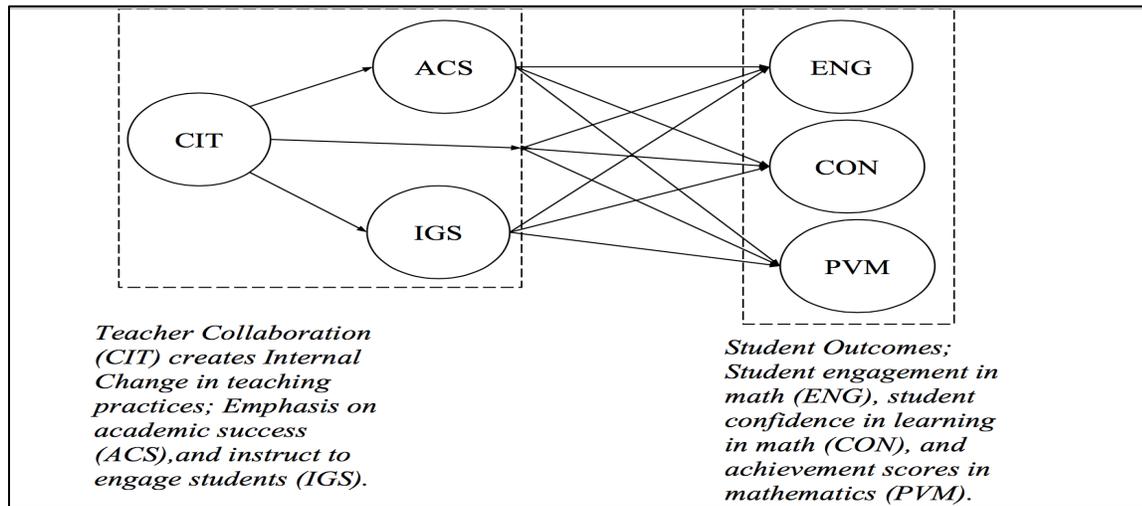


Figure 1. This diagram depicts the hypothesized direct and Indirect relationship between teacher collaboration, teaching practices (ACS & IGS), and student outcomes (ENG, CON, and PVM)

1.4 Study Problem

In recent years, the Saudi Arabia educational system has been undergone two major reform projects. The first occurred in 2007 when the government established and subsidized the largest educational reform project ever in Saudi Arabia called the King Abdullah Bin Abdul-Aziz Al Saud Project for Public Education Development (KAAPPED). This ten-year project has targeted four priority areas: furthering teacher skills, developing curriculum, enhancing extracurricular activities, and improving school environments (King Abdulla Bin Abdul-Aziz Project for Public Educational Development, 2011). The second project established the Education Evaluation Commission (EEC) in 2013 that is mainly responsible for accrediting K-12 schools and licensing educators. The tremendous budget that has been allocated to these reform projects, among others, has put Saudi Arabia among the top ten countries in the world spending the highest percentage of resources on education (Maroun, Samman, Moujaes, & Abouchakra, 2008). Almost 5.14% of the GDP, and 19.26% of the total Saudi government expenditure in

2008, was spent in educational projects (UNESCO, 2015). Yet, the performance of Saudi students as one significant educational indicator has not significantly improved (Alyami, 2014). Despite this generous funding, Saudi students have consistently performed below the average in international educational comparative studies. For example, during the last four cycles of the TIMSS (2003- 2015), Saudi students scored below average in math and science compared to peer nations both in the region and globally. Unfortunately, interpretations of this phenomenon are limited. One of the available interpretations is the conclusion reached by World Bank researchers; after analyzing educational reforms in the Middle East and North Africa (MENA) countries, they concluded in a report titled “The Road Not Traveled” that most of the attempted educational reforms in the region focused on “engineering changes” by which the school buildings, teachers, and curriculum were provided (World Bank, 2008, p. xv). Other potential interpretations can be obtained from the work of Wiseman, Alromi, Naif, and AlSadaawi (2008). After they investigated several educational indicators in Saudi Arabia and their impact on student achievement, they suggested that one key component to successful reform in the Saudi educational system could be through investigating school culture and its influence on other school components. Further, Dodeen, Abdelfattah, Shumrani, & Hilal (2012) suggested that teaching practices in Saudi schools should be more emphasized for better student outcomes. TIMSS (2011) will empower the current study to present further interpretations of the phenomenon by considering school culture in Saudi Arabia. Particularly, the study investigates teacher collaboration as a key component of school culture and its impact on improving teaching practices and ultimately students’ outcomes.

1.5 Study Hypotheses

Based on the conceptual framework, the study hypothesized the following:

1. Fourth-grade students' engagement, confidence in learning, and achievement score in mathematics significantly and positively are associated to their teachers' collaboration.
2. The association between fourth-grade students' engagement, confidence in learning, and achievement score in mathematics and their teachers' collaboration is mediated by their teachers' Emphasis on Academic Success (ACS), and Instruct to Engage Student in Learning (IGS).

1.6 Contribution to Literature

This study contributes to the literature in several ways. First, it extends the existing limited knowledge regarding the relationship between teacher collaboration and student outcomes. Although several studies have examined teacher collaboration in different ways (Goddard et al., 2007; Kelchtermans, 2006; Miller et al., 2010; Waldron & McLeskey, 2010), studies that empirically link teacher collaboration to student outcomes other than achievement scores are still limited (Goddard, Kim, & Miller, 2015; Ronfeldt et al., 2015; Schleicher, 2011). Beyond achievement scores of the main three subjects (mathematics, science, and reading), the study investigates two more student outcomes, student engagement and confidence in learning. Second, despite the scarcity of empirical studies in teacher collaboration and their impact on student outcomes, most of these studies were conducted in Western countries, particularly in the United States and United Kingdom. Therefore, this study expands the generalizability of the existing literature through examining teacher collaboration in a different culture, Saudi Arabia, in which this issue has not yet been empirically investigated (Wiseman et al., 2008). Third, and most important, as collaboration in schools becomes increasingly common in Western countries, particularly in the United States and Canada, most of the previous empirical studies that have linked teacher collaboration to student outcomes examined formal collaboration that was

externally administered to schools as a part of reform initiatives by school administrators, school districts, or educational departments (e.g., Goddard et al., 2015; Ronfeldt et al., 2015).

Alternatively, this study focuses specifically on informal teacher collaboration (Mullis, Martin, Gonzalez, & Chrostowski, 2004) as a professional learning opportunity initiated by teachers themselves through their interactions and reflections upon their practices.

1.7 Significance of the Study

In the Saudi educational context, this study was timely as it was conducted at a critical point between two educational reform projects. The first project is KAAPPED, which was just completed a decade after its implementation. As mentioned previously, this project has focused mostly in improving teacher quality “human capital,” and paid less attention to teaching quality through examining issues more related to “social capital” such as teacher collaboration and collegiality (Alnahdi & Arabia, 2014; Alyami, 2014; Wiseman et al., 2008). The second project is the ambitious and comprehensive vision that was just launched in 2016 by the Council of Ministers in Saudi Arabia, the highest legislation authority in the country. This Saudi 2030 vision was preceded by the National Transformation Program 2020 (NTP; Saudi Vision 2030, 2016). Both reforms included several educational, social, and economic strategic goals and initiatives. Educationally, the NTP includes eight strategic objectives. The second and the fourth objectives are the most important to this study. The aim of the second object is to improve recruiting, training, and developing teachers, and it focuses on professional development programs. The aim of the fourth object targets improving curricula and teaching methods. However, it was not clear how these objectives were going to be implemented. Despite the recent research findings that suggested focusing only on the teacher quality of human capital is not adequate to improve school effectiveness (Akiba, LeTendre, & Scribner 2007; Bryk, Camburn, & Louis, 1999; Liang,

Zhang, Huang, Shi, & Qiao, 2015), it is noteworthy that most of the key performance indicators of these two objectives have continued in approaching school reform through teacher quality “human capital” and paid less attention to improving teaching quality through social approaches, such as teacher collaboration (Alnahdi & Arabia, 2014; Alyami, 2014; Wiseman et al., 2008).

Therefore, while this study reflects on previous educational reform approaches in Saudi Arabia, it intends to guide educational reformers and policy makers in Saudi Arabia regarding the crucial role of teacher collaboration in improving teaching practice and school effectiveness. Furthermore, as one of the few studies of teacher collaboration that utilized TIMSS data, the results of this study may encourage other researchers to benefit from the participation of Saudi Arabia in the International Large Scale Assessment (ILSAs) studies such TIMSS, PIRLS, and PISA to investigate different variables in Saudi Arabian educational system.

1.8 Study Rational

In the Saudi educational system, there is an unprecedented level of importance on the ILSAs such as TIMSS. The Saudi 2030 Vision that just launched has focused on the educational part of student achievement scores in the TIMSS as substantial indicators in measuring student progress (see Appendix 1). Although achievement scores are widely used as indicators of educational systems, they have been criticized as they only explain part of the picture (Newmann, 1992). Therefore, a need exists for other measures of student characteristics that can explain student-learning mechanisms and contribute to improving their performance. Student engagement and confidence in learning are appropriate candidate measures of this mission (Wiseman, Alromi, Naif, & AlSadaawi, 2008). However, the literature suggests more rigorous investigations of constructs provided in TIMSS data (Marsh et al., 2013). Therefore, in order for these two measures to be effectively used within a multilevel data set such as what TIMSS

requires, their construct validity should be accurately assessed. The current study will apply the Multilevel Confirmatory Factor Analysis (MCFA) technique to evaluate the construct validity of a two measures, which are student engagement and confidence in learning in mathematics, for a sample of Saudi fourth grade students.

1.9 Overview of Dissertation

This study consists of five chapters. The first chapter provides an introduction to the dissertation including the theoretical framework that guides this study. It also includes the statement of the problem, the purpose of the study, as well as the specific research questions that undergird this dissertation. Chapter Two includes the literature review. Chapter Three addresses the methodology of the study. It includes sampling procedures, data collection, the study variables, and analysis procedures. Chapter Four reports the study findings. It includes measurement model, structural model, and the analysis software. Chapter five discusses the study findings. It also presents several implications of the study's findings, and suggests new directions for future research.

CHAPTER 2

THEORITICAL FRAMEWORK

2.1 Teacher Collaboration and Adaptive Expertise

In teacher professional development, there are three widely documented problems in learning to teach that need to be considered. The first problem is what Lortie (1975) called “the apprenticeship of observation,” which is that new teachers allow their preconceptions about teaching and learning while they were students at schools to affect how they teach (Darling-Hammond & Bransford, 2007, p. 359). Overcoming this problem requires teachers to “think like a teacher” and understand teaching differently from what they had observed when they were at schools as students, and try to apply conceptions they learned to teach (Darling-Hammond & Bransford, 2007, p. 359). The second issue is the problem of enactment, where teachers are required not only to understand what should they teach “knowing,” but also how they teach “doing,” and in most times handling this simultaneously. The third problem that limits teachers’ professionalism is the problem of complexity (Darling-Hammond & Bransford, 2007). While there are some aspects of teaching that can become routinized, there are many other aspects where this cannot happen because they are contingent upon student needs and unanticipated classroom events. Therefore, it is imperative for teachers to consider this complexity and develop metacognitive habits that would help them to think systematically about their decisions, which would help in their continuous development (Darling-Hammond & Bransford, 2007). Schwartz, Bransford, and Sears (2005) discuss three kinds of knowing: replicative, applicative, and interpretative knowing. Most of learning, as Schwartz et al. (2005) argue, occurs through the first two kinds of knowing, where people try to replicate or apply an old experience that they have

been learned from previous situation to solve a new problem in a new situation. Interpretative knowing, however, while important in a learning situation, is the most neglected type of knowing to learn. While the replicative and applicative types of knowing require regurgitation of the old experience to the new learning situation, interpretative type is about the interpretations that people have for the new situation and how these interpretations can help them to frame the new problems in a certain way in order to have an effect on subsequent thinking and learning processing. It is noteworthy, however, that emphasizing the interpretive approach does not mean neglecting the replicative and applicative approaches. Instead, it advocates an integrated approach to these various types of knowing (Schwartz & Bransford, 2005).

To improve teaching as a profession, it is important to realize that teaching knowledge does matter (Darling-Hammond & Bransford, 2007). Therefore, educators have concerns about how this knowledge, whether it is content or pedagogy knowledge, is built so that it is done in such a way that prepares teachers for “a whitewater world” that will impact educational goals and teaching strategies (Schwartz & Bransford, 2005, p. 42). Building on the idea of Schwartz and Bransford’s (2005) combination of the three types of knowing, which are replicative, applicative, and interpretative knowing, many educational programs are adapting the idea of “adaptive expertise.” This “gold standard for being a professional” differentiates between two dimensions of expertise, efficiency and innovation as shown in Figure 2 (Darling-Hammond & Bransford, 2007, p. 67). In the efficiency dimension, teachers, particularly novices are more likely to ask about “how to” techniques that would help them perform their teaching duties efficiently. They build routine expertise that enables them to solve many future problems efficiently depending on the replicative and applicative solution they had encountered in previous situations. Although this kind of pure efficiency is crucial in early years of teaching, it

is less likely to establish lifelong learning that requires more techniques of teaching, and that is where the second dimension of expertise, innovation, comes to play (p. 77). This dimension requires the ability to unlearn previous routine expertise and let go of beliefs and initial ideas that had been used to frame previous problems in the first place. Resisting these beliefs and assumptions is vital for what Schwartz and Bransford (2005) called “the sudden cessation of stupidity” that constrains thinking to solve a problem within a limited space and tolerates the ambiguity of having to reconsider and refine old preconceptions and frames (p. 45). This restructuring of previous core ideas and beliefs is neither preferable nor easy, and it may seem to reduce the other dimension, efficiency, in the short run. However, it will enhance the total affect of adaptive expertise in the long run. Researchers have asserted that these two dimensions of adaptive expertise should not be seen as mutually exclusive. Instead, there should be a balance between both of them to keep this expertise in the “optimal adaptability corridor” as shown in Figure 2 (Schwartz & Bransford, 2005, p. 55).

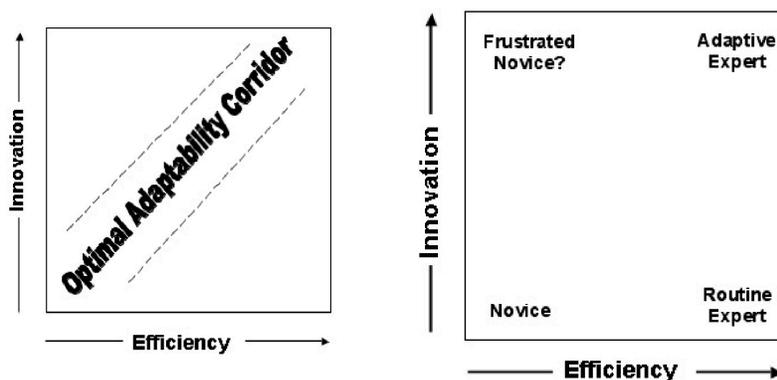


Figure 2. Two dimensions of learning and transfer: Innovation and efficiency and the optimal adaptability corridor. Adapted from Schwartz & Bransford, 2005.

The current study argues that a teaching approach that balance these two dimensions to be within the corridor of adaptive expertise can provide opportunities of actively interaction and collaboration between colleagues so they can elaborately reconsider their previously held beliefs and conceptions about teaching and thus enable them to reframe new teaching approaches that may improve their teaching and thereby benefit their students.

2.2 School Contexts and Teachers' Learning Opportunities

Historically, teacher workforce quality, beyond the initial hiring and routine staff evaluation, was not seen as a direct responsibility of public schools. As Johnson described it, "Schools are in the business of promoting students' learning and growthIronically, public schools are not in the business of promoting teachers' learning and growth" (As cited in Darling-Hammond & Sykes, 2009, p. 257). With the escalation of school effectiveness initiatives that have focused on the school as a unit of change and with research evidence that a " professional staff will work toward implementing strategies and programs to improve results," teacher learning becomes, instead of a luxury, a necessary task of schools (Hord, 1997, p. 7). Subsequently, during the last two decades, there was a shift in school reform literature that has proposed viewing schools as professional learning communities rather than organizations (Hord, 1997). The influential work of Senge (1990) describes a continuous learning organization, yet professionalizing schools has been seen as a daunting task. This challenge resides in the complexity of the nature of teaching that requires extraordinary teachers' efforts. The learning demands in this kind of work cannot be fully anticipated or satisfied by the pre-service programs (Little et al., 2009). Therefore, workplace contexts of teaching can deeply enhance or diminish opportunities and resources of teachers' learning. For example, the lone wolf scenario, where

teachers are left to sink-or-swim is less likely to support teachers' lifelong learning or establish their professionalism. As has been sociologically established, complex tasks require developing robust lateral relations. As one of school leaders described it, "One could imagine schools in which teachers are in frequent conversation with each other about their work, have easy and necessary access to each others' classrooms, take it for granted that they should comment on each others' work, and have time to develop common standards for student work" (Darling-Hammond & Sykes, 1999p. 233). Although research in how school context affects teaching and teachers learning did not provide a simple checklist nor a template that can be translated into a specific educational policy, it does embrace "a set of obligations, opportunities, and resources for teacher learning" (Darling-Hammond & Sykes, 1990, p. 257). These opportunities for teacher learning have been widely researched. They could be categorized into two sets: structural and cultural resources (Little, 1990). Aspects that constitute school structural resources that might foster or hinder teacher learning and growth through their collaborative activities could include shared responsibility among school community, school size, teacher assignment, out-of-classroom teacher time, and classroom visitations. Each of these conditions will be discussed.

Shared responsibility among school community has intensively investigated in school effectiveness literature (Dufour, 2004; Hord, 1997; Louis & Kruse, 1995; Stoll, Bolam, McMahon, Wallace, & Thomas, 2006) . For example, in his informative chapter "Organizing School for Teacher Learning," Little (1999) suggests that shared responsibility is a structure that creates a genuine interdependency among the school community that contributes to one another's success with students (Darling-Hammond & Sykes, 1999). Noting the importance of this concept in school success, Little (1990) asserts that establishing shared responsibility among school community is a challenge. In trying to foster this collective commitment, school leaders

encounter tension between teachers' personal autonomy and common commitments. In some cases, this tension is not resolved properly. To illustrate, while some formal arrangements of schools show paths of teacher collaboration and organized shared responsibility, teachers' work generally remained unchanged. Therefore, researchers asserted that the power of shared responsibility for students is demonstrated in schools where structure is enforced by culture (Peterson, Brietzke, 1994; Darling-Hammond, Sykes, 1999; Hord, 1997; Kelchtermans, 2006).

School size is another context that has an impact on teacher learning. While there is mixed data about the benefits of small sizes of schools and how that size can facilitate teacher growth, large size schools could exacerbate teacher problem learning in different ways. Little (1999), for instance, mentions that large scale makes it more likely for teachers to look closely and excuse themselves from shared responsibility for the same students, often in the same subject and grade. Other effects of large scale is the anonymity that is created between teachers, making it less likely for them to get know each other personally and professionally in a way that could contribute to their mutual learning. Another potential element that could affect teachers' growth is the sheer number of out-of-class activities in large schools that puts teachers at a disadvantage with their joint-work.

Teacher assignment is another school context in which school policies and practices could join to strengthen or diminish teacher-learning opportunities, and therefore teaching quality (Darling-Hammond & Sykes, 1999). Little (1999) delineated three complementary criteria to use when judging teacher assignments. First is the assignment fit that means making the most of the existing teacher's knowledge, experience, and interest at the individual classroom level. A plethora studies suggest that teacher confidence in the content or group that she or he teaches is a crucial for teaching quality, and hence student learning (Caponera & Losito, 2016;

Johnson, Berg, & Donaldson, 2005; Peterson, Brietzke, 1994; Reeves, Pun, & Chung, 2015; Williams, 2010).

Ironically, teacher seniority and internal school policies continue to play a vital role in teacher assignments in many schools (Darling-Hammond & Sykes, 1999). The second measure of appropriate teacher assignment is what Little called “stretch” (1999, p. 241). This criterion underlies the idea that teacher’s assignments should consider not only the assignment good fit but also need to extend more opportunities for teachers to grow professionally. It is noteworthy that this criterion differs from school to school. Therefore, it is a school community task to discover the best “stretch” situations for its team members and act to maximize new learning and count as an existing teacher experience.

The third teacher assignment criterion is the configuration of teachers’ assignment in regard to their collegiality relationship. When a school assignment is considered a good fit and provides more stretch for teachers, it is likely to create a community of practitioners that reinforce the collaboration instead of competition among colleagues (As cited in Darling-Hammond & Sykes, 1999).

Out-of-classroom teacher time is the fourth vital school element that can influence teacher-learning opportunities. During this time, teachers can examine their ideas and reflect in their practices, which contribute to each other’s teaching success. Although there is no golden standard for allocating time that might be sufficient, researchers suggested that the ratio of in-class to out-of-class time during the scheduled school day might be considered a measure of this time (Little, 1999). Recently, this factor has been in the limelight from researchers globally, as they are stimulated by its effects on teaching quality that are conducive to better student outcomes. For example, Darling-Hammond and Wei (2009) in their widely cited report

“Professional Learning in the Learning Profession” found that by contrast, U.S. teachers spend about 80 % of their total working time engaged in classroom instruction, while other nations’ teachers spent only about 60 % of their working time in classrooms, which provides them much more time to plan and learn together. Admittedly, out-of-classroom teacher time has encountered several challenges. One of them is the absence of the physical space in school buildings that teachers can utilize to meet and practice their professional activities. Ironically, while studies present evidence that out-of-classroom teacher time and its logistic services such spaces and materials are crucial for a rich learning environment for both teachers and students, these physical spaces are rare in school buildings, even the recently structured ones (Darling-Hammond & Sykes, 1999). This challenge is more likely to result from another issue. In the era of increased accountability, school administrators and other interested groups exercise more pressure for concrete indicators of productivity. As school policies often do, framing out-of-classroom teacher time with constraints inhibits rather than fosters teachers’ learning by treating this time’s activities as if they were a short-term byproduct rather than an ongoing process of inquiry (Darling-Hammond & Sykes, 1999).

A fifth school context that can influence teacher learning is the availability of classroom and school visitations. Glickman, Gordon and Ross-Gordon (2001), called this practice “deprivatization of teaching” by which they mean that teachers’ classrooms are open so that teachers can observe each other (p. 470). This practice might contribute to teachers’ development in two different ways. First, it ignites new ideas in different aspects of teaching, especially when teachers are able to visit several classrooms on different occasions. Second, it provides teachers the opportunity for fruitful discussions, reflective dialogue, and inquiries about students, teaching, learning, and challenges that they face and how they can be managed. Ironically,

although the crucial role of these observations in teachers' development has been noted (Holmes, Futrell, Christie, and Cushman, 1995; Little & et. al, 1987;), they remain rare practices (Lomos, Hofman, & Bosker, 2011). This rareness can be attributed to the traditional norm of teaching as a private task, where any ambitious effort of teacher collaboration stops at the classroom door. However, the vital role of deprivatization teaching practices in teachers' learning and development, as aforementioned, requires more investigations of teachers' reluctance toward a balance of their autonomy in classrooms and openness towards their peers.

The other set of teacher learning opportunities that schools provide teachers, beside school structural aspects, is school culture. Studies assert that structural supports for teacher learning alone will not succeed in causing teacher learning without compatible values, beliefs, assumptions, and traditional norms that support learning opportunities among teachers (Hord, 1997; Little, 1990). Peterson, McCarthy, and Elmore (1996) conclude from their three case studies that "school structure follows from good practice [culture, and] not vice versa" (pp. 148-149). While researchers define school culture or ethos in different ways, they have something in common. This commonality in the definition rests on invisible, taken-for-granted flow of beliefs built gradually by the school community that supports its development. In another deliberative definition, Little for example, depicted school culture as the "woven fabric" of daily schoolwork that teachers among a school community form gradually during their professional life (1999, p. 253). In another definition, Peterson & Brietzke (1994) defines school culture as "A complex web of norms, values, beliefs and assumptions, and traditions and rituals that have been built up over time as teachers, students, parents, and administrators work together" (p.6). They continued to distinguish between different types of school culture and focused on a collaborative school

culture in which “underlying values and beliefs reinforce and support high levels of collegiality, team work, and dialogue about problems of practice” (p. 7).

2.3 Collaborative School Culture

Three decades ago, Lieberman (1988) made the case that schooling needs not only an elite cadre, but it needs professionals in the educational system. He argued that this professionalism would be costly. However, it is needed not only to attract talented people to the profession, but most importantly, it is needed for the student’s learning. (p. 9). Lieberman (1988) posted that new economic and demographic demands that create new nature of jobs are most likely to leave youth unemployed. Therefore, rhetoric about critical thinking by students and problem solving in the curriculum has disappeared. However, this rhetoric has not been translated into action. The interpretation of this inaction is not only due to conceptions of teaching and learning in educational system, but also because these skills are not rewarded from the large society. Developing these skills requires thinking and insights on how teachers should deliberately organize and conduct their work in a way that encourages their students to do the work of learning.

Cognitive research provides several sets that illustrate the need for a professional culture in schools. One of these sets that deal with cultivating problem solving in classrooms is the complexity of schoolwork. Proceeding to teach a certain concept in a classroom requires juggling many cognitive aspects simultaneously. It is not only a matter of delivery of the concept, but also and most importantly, a complex task of structuring that concept by students themselves where teaching is more than delivering the knowledge and the learning is more than acquisition of knowledge. Another set is distinguishing between what Lieberman called “well-structured problems and ill-structured problems” (Lieberman, 1988, p. 15). In the past, most of the

problems that teachers dealt with were well-structured problems for which they had already have received training. However, when imagining schoolrooms as workplace and students as intellectual workers, teachers and students alike are more likely to experience ill-structured problems that are not clearly stated nor fit for a single answer through which collaborative learning can take place and becomes the learning style of the school environment.

Cognitive research offers other findings that illustrate the difficulties of cultivating critical thinking in classrooms and the need to develop a professional culture in schools. Out of these findings is the number of students in the classroom. The above sections have showed how difficult it is for teachers to communicate a certain concept. Imagine this situation not only with a small number of learners but also in a classroom of 25 or 30 students simultaneously. An accountability system is another issue that obstructs teachers and their students alike from practicing critical thinking skills. There is an implicit contract between teachers and students that involves exchanging performance for grades (Lieberman, 1988). This approach encourages avoiding ambiguity and ill-structured problems in teaching contexts, by which students are more likely to submit precise answers to get good grades. In his words, Lieberman (1988) describes this contract pointing out that “The accountability system exerts a powerful influence on teachers and students alike, pulling both away from the sorts of tasks that are complex and difficult to assess.” (p. 18). To summarize, professionalizing schools in which teaching is constructed as leadership and learning as production not consumption is more likely to encourage both teachers and students to realistically practice a higher-order thinking agenda.

The look into the teaching as a career or job is another vital factor in cultivating a collaborative culture. Lieberman and Miller (1988) distinguish between teaching as a career or a job. They argued that two characteristics delineate the line between these two sets. These

characteristics are the level of opportunity and the level of capacity in teaching, and they note that teaching is a career asking for high levels in both characteristics. They suggested several features of school environment that support teaching as career. A collegial environment was one of the critical features of supportive schools. These kinds of relationships in school environments enhance both factors that researchers proposed for a teaching as a career, which are the level of opportunity and level of capacity. These developmental relationships do not have to be formally designed. Indeed, researchers proposed that voluntary and naturally formed collegial relationships might be more effective than ones that are officially organized (Lieberman & Miller, 1988; Little, 1982). In their study that surveyed teachers about the importance of informal conversations with colleagues comparing to other formal professional development, Lieberman and Miller (1988) found that 78% of the sample indicated that informal conversations with colleagues were either very helpful or somewhat helpful. In contrast, other professional development resources, whether universities courses or school district sponsored activities, were noted by 22%, and 8% respectively as being very helpful. These kinds of relationships differentiate effective schools from less effective ones. While the isolation, egg-crate model is the norm in less effective schools; interactions between teachers are the norms in most effective schools. These interactions have been described by one of the teachers as:

I want nonthreatening feedback from someone who has the time to really take a hard look. It would have to be someone whom I respected and looked up to, and they would have to value the same things I do in teaching. I need a comfort zone, a framework around me within which I have the freedom to be myself- to use my own judgment and get trust and respect. (Lieberman, 1988, p. 35)

In the conclusion of their chapter, making the case for professionalism in schools, Devaney and Sykes (1988) underscored that the way to build a professional culture in schools is not impossible, but it is tough. It requires hard work from “the inside, not mandates from outside, by knowledgeable, experienced, thoughtful, committed, and energetic workers- teachers” (p. 20). These change agents, teachers, are less likely to be enlisted from the outside for two main reasons. First, for change to occur needs enough committed numbers of its agents, and it is not likely to have these numbers from the outside. Second, as Devaney and Sykes put it, “social cultures cannot be copied--they must grow themselves” (p. 20). Therefore, a school’s professional culture should be importantly and firmly built on the learning and teaching agenda needed for that culture. It is crucial, then, for schools in their efforts to build a professional culture that they not only identify and recognize but also stimulate their teachers to move toward professionalism.

An ample evidence exist that shows collegiality is an essential prerequisite for building a more professional school culture. Little (1987) found that teacher collaboration, not isolation, was the norm in the more professional schools. Little’s finding was supported later by Rosenholtz (1989) when she distinguished between two kinds of school culture, collaborative and isolated schools, based on the level of the collegiality work that is led usually by teacher-leaders. Other researchers suggested several skills in different roles that these teacher-leaders should manifest to create a collegial culture in their schools (Lieberman, Saxl, & Miles, 1988). Among these skills are building trust, rapport, and confidence in others. The researchers delineated several key elements found in these collegial cultures. One of these elements is that teachers in these environments depend on each other for their professional development, believing that expertise exists mostly within, not out of, their workplaces. They assume that

professional knowledge is inherently situated in their day-to-day work and can be best captured through reflecting and sharing the same experience (Vescio, Ross, & Adams, 2008). Another element is that they see themselves as not only practitioners, but also theoreticians, where they can create new knowledge and use it.

Recently, this later element has been supported through what are called Professional Learning Communities (PLCs), one of the key contributions of the sociology in school culture and change (Leithwood, Seashore, Anderson, & Wahlstrom, 2004). Despite the ambiguity of this concept, PLCs' (Dufour, 2004) central idea in developing school culture emphasizes professionalism relies on establishing a school-wide culture that “makes collaboration expected, inclusive, genuine, ongoing and focused on critically examining practice to improve student outcomes” (Leithwood et al., 2004, p. 66). The main idea of establishing this school-wide culture in this model of school reform is that what teachers accomplish together outside of classrooms can be as critical as what they can accomplish inside in developing their practices and students' learning (Leithwood et al., 2004; Little, 1993). It is noteworthy, however, that these reflective and collaborative activities are not the goal. Instead, they should be a procedural way of creating environment of “thinking, deliberation, and experimentations” that develop a change orientated mode of teacher works (Richardson, 1998, p. 5).

2.4 Teacher Change

Teacher collaboration is considered a central element in facilitating school change, restructuring, and teacher professionalism. However, there has been an argument around this issue. First, where does that change should come from; voluntary or forced? Second, how does this change balance individual autonomy with the collective responsibility of teachers? Supporting the internal change, Morimoto (1973) argues, “When change is advocated or

demanded by another person, we feel threatened, defensive, and perhaps rushed. We are then without the freedom and the time to understand and to affirm the new learning as something desirable, and as something of our own choosing. Pressure to change, without an opportunity for exploration and choice, seldom results in experiences of joy and excitement in learning” (p. 255). This idea unfortunately not only ignored by those who advocate external change, but it also misinterpreted to read that teachers do not want to change (Richardson, 1998). In the same line, Richardson (1994) in her long-term collaborative study of teacher change found that teachers do really engage in voluntary change to support their students’ needs and what works in their contexts. Moreover, she underscored that while these changes are minor, they can be dramatic. On the other side of the argument, Cuban (1993) argued that these teachers’ voluntary changes are minor and not adequate to enhance their learning. Other researchers found that teachers who make a voluntary change may decide to make some changes that are not based on supported assumptions (Richardson, Anders, Tidwell & Lloyd, 1991). Therefore, they suggest guidance of these voluntary changes so that their underlying beliefs and assumptions could be examined. The questions then would be to which extent should these guidance efforts be offered? How could the balance between individual autonomy of teachers and their collective responsibility be achieved? While this balance is extensively described in the literature particularly in the PLCs movement reform, in reality it is difficult to achieve. In schools where there are continual critical discussions about aims, standards, and procedures, researchers suggest schools should be considered as communities of practice in which individual autonomy is considered within these communities (Pendlebury, 1990; Richardson, 1998). Little (1992) called this sense of autonomy and responsibility that goes beyond individuals to school level civic responsibility. To the extent that civic responsibility should be approached to facilitate teacher change, however, could be

critical issue. Little (1992) caution against what he described as "formally orchestrated" collaboration where teacher change efforts would be bureaucratic and contrived. This contrived collegiality is enhances administrative control rather than facilitates collaborative culture in schools.

While formal teacher collaboration has been investigated in the literature, informal collaboration has not been adequately empirically examined (Akiba, 2012). Most studies that investigate informal teacher collaboration are qualitative in nature. Scribner (1999) investigated in a multi-site case study of various kinds of professional development most teachers are involved in, the most beneficial professional development (PD) activities, and the reasons they were involved in those types of PD. Among the informal learning opportunities, teachers were involved in individual inquiries and collaboration. For professional development, the largest intrinsic motivation factor was teachers' desires to improve their content knowledge. In an intensive case study, a team of researchers tracked 13 teachers through their diaries of their learning activities for seven days on two occasions, first in the late 1990s and then in early 2000 (Smaller, Hart, Clark & Livingstone, 1997). These diaries were then followed by interviews. After analyzing the data, they found that teachers spent about seven hours a week on informal learning including meetings and conversations. In a more comprehensive study, Lohman (2005) surveyed 318 public school teachers and human resource development professionals regarding their informal learning activities. The study found that teachers relied on the feedback and expertise of others through collaboration, observation, and informal conversations. While these qualitative studies expand the knowledge about these informal learning opportunities for teachers, further quantitative studies are needed to investigate these informal learning opportunities of teachers, specifically collaboration, and its impact on teaching practices and

student outcomes. Based upon the studies above, it can be stated that collaboration can be defined as joint interaction in the group in all activities that are needed to perform a shared task. This concept is not static and uniform, but different types of collaboration can occur with varying depths. In essence, collaboration can be seen as an umbrella term of several collaborative activities.

2.5 Student Engagement

TIMSS fundamentally is an achievement study. Therefore, most studies that utilized this dataset included achievement scores as a main outcome variable (e.g. Martin and Mullis, 2011; House & Telese, 2015; Abu-Hilal, Abdelfattah, Shumrani, Dodeen, Abduljabber, & Marsh, 2014). Beyond the achievement scores, the current study will include student engagement and their confidence in learning as other student outcome variables. Student engagement and confidence in learning are two critical concepts to educational success. Students who engage and have more confidence in their learning are more likely to show interest, expend effort, and persist through learning difficulties. These two concepts have been fundamental in the effective schools research since the 1970s (Newmann, 1992). In this section, both variables and their relationship to instructional and learning contexts will be discussed.

Engagement in learning is a construct that refers to “student's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills or crafts that academic work is intended to promote” (Newmann, 1992, p. 12). Literature distinguishes between three types of this construct: effective, behavioral, and cognitive (Fredricks, Blumenfeld, & Paris, 2004; Mullis, 2011; Cai, & Liem, 2017). While the effective engagement represents students’ feelings during their learning process, behavioral engagement indicates students’ efforts and involvement in tasks and cognitive engagement refers to thinking strategies

(e.g., elaborating, comparing) students use in their learning. Together, these three types of engagement are critically important for academic success. Recently Yu et al. (2017) underscored empirically how these three types of student engagement affect each other and how they together affect positive academic achievement in mathematics. While student engagement is widely studied in the literature, there is still exists overlapping between its elements and other constructs such as motivation. In their comprehensive review, Fredricks, Blumenfeld, and Paris (2004) described student engagement as an inclusive, multidimensional, and malleable construct. This later characteristic, malleable, is because it linked in the literature to several aspects including teacher support, classroom structure, autonomy support, and task characteristics. This inclusiveness of the student engagement concept makes it difficult to be precisely measured. Indeed, the challenge is twofold. If the aim is to capture every detail of this construct in its three dimensions, then it will not be practical because of constraints of time and resource. The other challenge in measuring student engagement is the overlapping that usually happens between its types. For example, as Fredricks et al. (2004) noted, questions about persistence and preference for hard work are included in measures of both behavioral engagement and cognitive engagement. Despite these challenges, student engagement has been considered a key factor for academic success (Cai & Liem, 2017). In this current study, two engagement measures are examined. The first measure investigates teachers' support for engaging their students during their teaching. This measure provides insight into the degree to which engaging student can explain variations in school and classroom changes. The second measure is the student-report based measure. Both measures assume to measure general engagement covering social (emotional and behavioral) and academic or cognitive aspects.

2.6 Student Confidence in Learning

Since the 1970s, this concept has received the greatest amount of attention from social science researchers. When examining other variables, self-confidence is positively correlated variables to student achievement, particularly in mathematics (Kloosterman, 1988). Furthermore, students who are confident in learning mathematics are more likely to take mathematics when it is optional and are more comfortable dealing with mathematical challenges (Kloosterman, 1988). Psychological literature distinguishes between two inherent concepts of confidence in learning. These concepts are the academic (for example, the mathematics and science self-concept) and the non-academic components of the self-concept or confidence. Kung (2009) showed that while academic achievement is substantially related to the academic self-concept (domain-based), it is almost unrelated to the non-academic components of the self-concept.

Both engagement and confidence imply two noteworthy aspects. First, they are not readily observable characteristics. Rather, they are latent constructs that describe a variety of students' characteristics toward learning. The other essential aspect is the continuum of these two constructs, as engagement and confidence cannot be viewed as a dichotomous state. Instead, they are viewed in a continuous state from less to more, which implies different levels of these constructs (Newmann, 1992).

2.7 Educational Reform Profile in Saudi Arabia

The Saudi educational system was established relatively recently. The Ministry of Education was instituted in 1954. Since then, it has been tasked with supervising public and private education sectors including primary, intermediate and secondary schools. Since its inception, the Ministry of Education has exerted efforts to educate the nation's youth. Among these efforts was the reduction of illiteracy among citizens. While illiterates were estimated to

constitute more than 90% of the Saudi population in 1950, this percentage was significantly decreased to constitute less than 14% by the end of 2011 (UNESCO-IBE, 2011). Other accomplishments can be observed through the gross enrollment ratios across the educational systems. In primary education, for example, the ratio increased from 82% in 1990 to 99% in 2009 (UNESCO-IBE, 2011). In the same year, 97.2% of first graders reached fifth grade. In other example, 88.7% of boys and 89.6 % of girls of intermediate schools reached the first grade of secondary school by 1995-96. Secondary schools have been expanded as well. From 2000-2010, secondary schools increased by more than 50%. In particular, in 1990-2000, there were 2,938 secondary schools, more than half of them for girls. Ten years later, these numbers increased to 4909 schools half of them for girls. In the same period, the total enrolment of students was 1,096,174, of whom 490,112 were girls taught by 99,753 teachers and 4,756 administrative staff (UNESCO-IBE, 2011).

Despite these efforts to expand schooling, pressure from families and civic leaders persisted on the Ministry of Education to improve not only the quantity, but also the quality of the educational system. In response to these pressures, three vital reform efforts were initiated during the 1990s. First, a new cadre of education staff was approved. This cadre granted teachers, administrators, and other school related jobs a 20-30% higher salary than their corresponding government employees (UNESCO-IBE, 2011). Corresponding with this cadre, teachers' qualifications have been raised to be at least a bachelor degree obtained from a four-year teacher college or schools of education instead of a diploma from a two-year institute. In addition, maternity vacation for female teachers has been extended to two fully paid months. The second reform targeted high schools and their failure to achieve its main goals, whether in preparing students for higher education or the marketplace. This reform aimed to improve the

traditional high school policies and practices in the country. It embraced, for example, the credit hour approach where there was no common regular school schedule for all students, and students are allowed to work part time while they are pursuing this type of school (Alotabi, 2008). The reform offers more than six different tracks, instead of only two in the traditional high school, and it required 150 credit hours for graduation. The third reform was the Comprehensive Evaluation Program (CEP). According to the Ministry of Education, this program aimed to recognize the extent to which the school contributes to raising the level of achievement for students and to develop the students' patterns of behaviors and skills required by the educational policy in the state. The project prepared several evaluation tools including diagnostic tests to evaluate students in two subjects, Arabic language and mathematics; identifying learning difficulties faced by students in these two subjects; and preparing suitable remedial programs that might help schools to overcome these difficulties including a variety of teaching methods, alternative evaluations and assessment approaches, and teacher training. This program was planned into three phases. The first phase was implemented in grade one through grade three, the second phase included grades four through six, and the third phase was for intermediate schools. Unfortunately, except the educators' cadre, these three reforms that had shaped the Saudi educational system during the 1990s were not fully implemented, and because of the many critics and pressure mostly from families, these reforms came to the end.

Despite these reform efforts, educational quality remains a widespread concern for both families and government officials. One indicator for that ineffectiveness was the achievement level of the eighth-grade students as they participated for the first time in the TIMSS-2003. The achievement scores for both boys and girls were far behind the international average (Mullis et al., 2004). In a response to these disappointing results, a new educational reform was initiated.

In 2003, the Ministry of Education adapted a Ten-Year Strategic Plan to take effect from 2004-2014. This plan had included several goals for different levels of educational system. Among student outcome objectives, for example, this plan had stated that by the end of the program:

Students will have acquired practical knowledge, skills, and attitudes; they will be able to positively react to and face modern changes; they will be able to apply advanced technologies with efficiency and flexibility and to deal with international competition in scientific and practical fields. Their positive participation in an efficient educational system will allow them to develop appropriate abilities and attitudes and to spread the positive spirit of work at school environments that encourage learning and social education. (UNESCO-IBE, 2011, p. 1)

One of this plan's initiatives was establishing educational councils in each of the regional and provincial education bodies across the country. In 2006, these educational bodies had reached a total of 42, including 13 General Education Departments and 29 Provincial Education Administrations. The reason for creating these councils was to delegate operational and administrative responsibilities to the local educational bodies by which the Ministry of Education can focus on strategic issues. These councils aimed to democratize the decision-making process in education, reinforce the ties between educators at various levels, and open the doors towards society (UNESCO-IBE, 2011). Unfortunately, the implementation of this strategic plan, including the Educational Councils, has not gone far, particularly in its intent to institutionalize reform efforts. Observers have attributed these reforms' ineffectiveness to various reasons. Alsadaawi (2010), for example, suggested that most Saudi reform efforts were lacking standards that can be used to judge these initiatives. Other observers asserted that these initiatives were not planned at the institutional level, which put them in jeopardy of marginalization or even

cancellation (Al Issa, 2009). An example of the instability is that during the suggested period of the strategic plan (2004-2014), three ministers have led the Ministry of Education. During the last decade, many observers were skeptical about the capacity of the Ministry to juggle the administrative and operational work while initiating more reforms (Al Issa, 2009; Alsadaawi, 2010; Smith & Abouammoh, 2013). These skeptical views have been supported by the fertility rate that was among the highest rates internationally (2.1%) (World Bank, 2011), where a third of the population (5 million) go to more than 30,000 schools (Aburizaizah, Kim, & Fuller, 2016). Yet, the country faced unprecedented challenges. One was that the relationship between education and economic growth has remained weak. As a result, in 2009, Saudis constituted only 10% of the workforce in private sector in which jobs require highly qualified employees (Koyame-Marsh, 2017). Another challenge was the globalization that has led to a demand for a variety of skills and competencies that inevitably will influence the content and nature of what education systems should provide. Yet, the considerable resources that have invested in educational system in general and reform initiatives in particular have not fully delivered on their promises. During the period from 1980 to 2005, the educational budget in Saudi Arabia did not fall below 6.9% of its general domestic product (GDP). Yet, in 2004, it was ranked 97 out of 125 countries in the Educational Development Index (EDI) published by the United Nation (Maroun, Samman, Moujaes, & Abouchakra, 2008).

Consequently, the Saudi government decided to try a new approach of educational reform by creating two different and separate identities out of the Ministry of Education to lead the educational reform. The first project was the King Abdullah Bin Abdul-Aziz Public Education Development Project (KAAPEDP) (TATWEER) that was established in 2007. Six years later, the second project was initiated, which was the Public Education Evaluation Commission

(PEEC). These two initiatives are the latest in the Saudi educational reform scene. Both aim to reform educational system; however, they are different in their approach. The first project, KAAPEDP, was been initiated in 2007 by King Abdullah (1924-2015) and subsidized with a seven-year budget, the largest ever, at \$2.4billion (Alreshidi, 2016). While this project aimed mainly to reform the educational system, it was catalyzed by other economic crises in the country. The unemployment level among Saudis, for example, reached 12% in 2007. Ironically 55% of the total workforce in the private sector were non-Saudis (Maroun et al., 2008). These economic concerns have been combined with the aforementioned concerns of the capacity of the Ministry of Education to lead more reform projects, which has led to a holding company owned by the Public Investment Fund to administer the KAAPEDP. This business-based model intends to form partnerships with international companies to develop an educational industry in four priority areas: curricula, teacher skills, technology, and school environment (Alreshidi, 2016). In 2009, for example, KAAPEDP partnered with the U.S. publishing company McGraw-Hill to adapt a series of mathematics and science textbooks to the Saudi context. In its vision, the KAAPEDP states the following:

1. The learner is the focal point of the learning process: working to achieve excellence in learning for all learners, according to their abilities.
2. The Ministry of Education's role is to focus on educational planning, guiding the educational process, development of educational standards, and building quality and motivation systems.
3. Decentralizing the educational process administration and giving more authorities to educational regions and schools.

4. Building capacity and equipment in schools to develop the educational process and direct all its plans and programs to improve learning.

5. Building human and technical capacities at educational regions to guide the development process at their schools and achieve high quality performance (TATWEER, 2016).

Currently, KAAPEDP's holding company has created four small companies for educational services, transportations, buildings, and technology. Although this project is passing its first decade, data, particularly local indicators, remain scarce when it comes to evaluating its goals. Internationally, indicators suggested that no substantial improvements have been achieved yet. The Educational Development Index (EDI) by UNESCO, for example, ranked Saudi Arabia 66 among 115 participant countries. This index composite is made of four indicators: primary net enrollment ratio, adult literacy rate, gender specific index, and survival rate to grade 5 (UNESCO, 2011). Moreover, Saudi students' performance in the recent TIMSS-2015 in both grades and subjects (fourth, eighth, /math, science) was not only below the international average but also regressed from where it was in 2011.

In response to the unsatisfactory performance of the educational system, policy makers and educators realized that there was a fundamental lack of standards that should be used to judge any educational reforms (Alsadaawi, 2010). Consequently, the Public Education Evaluation Commission (PEEC) was established in 2013. The PEEC is independent from the Ministry of Education and the other educational bodies. It has been charged with four main tasks, including establishing an evaluation system of K-12 education, building qualifications for educators and licensing them periodically, building national standards for curricula across the K-12 education, and preparing and implementing national tests. The PEEC spent its first three years

establishing its systems. By the end of 2015, the PEEC implemented its first national tests. This first cycle aimed to measure student performance in mathematics and science for third and sixth grades. In this cycle, 25,000 students within 42 cities were tested. Beside these national tests, 1686 teachers within 562 schools and parents were surveyed. The second cycle of these diagnosis tests was implemented in April 2016. In this version, 10,000 students in the third and sixth grades within 46 cities were tested in the Arabic language. The PEEC plans to implement these national tests gradually and individually in different subjects during the next four years (PEEC, 2016). By the end of the first phase in 2020, and for the first time, the PEEC will test students in third, sixth, and ninth grades in the Arabic language, mathematics, and science.

Based on the findings of these two cycles of national tests, 40% of student performed below the grade level in both subjects. Notably, girls performed better in both subjects and grades as well. Moreover, there were notable differences between different provinces across the country. In addition, the surveys indicated that 40% of teachers did not participate in any kind of collaboration with other colleagues, mostly because of work conditions (PEEC, 2016). At this point, many observers consider the PEEC's efforts a cornerstone in educational reform in Saudi Arabia and hope that within the next few years these efforts through the cooperation with the Ministry of Education and other educational bodies will improve educational performance both nationally and internationally.

Recently, and in an unprecedented institutional reform, Saudi Arabia launched its 2030's Vision early 2016. The three main themes of this vision are a vibrant society, a thriving economy, and an ambitious nation. Although the main goal of this vision is to diversify the country's GDP by freeing it from extreme dependency on oil (Saudi's Vision 2030, 2016), its Transformation Program 2002 has a variety of initiatives within different government agencies

and bodies. Within the Ministry of Education’s initiatives, there are eight strategies. Among these strategies are improving recruitment, training and development of teachers, improving curricula and teaching pedagogies, and improving school environments. In fact, these strategies are not unique from what this Ministry has tried to achieve during the last two decades. The unique feature, however, is the accountability level of these strategies. For example, in the fourth strategy that targets to improve curricula and teaching pedagogies, there are four key performance indicators (KPIs).

Table 1

Strategic Objective 4 of the National Transforming Program (NTP), One of the 2030's Saudi Vision Programs

Strategic Objective (4)		Improve curricula and teaching methods			
Relevant Vision 2030 Objectives		Establish positive values and build an independent personality for citizens Provide citizens with knowledge and skills to meet the future needs of the labor market			
Key Performance Indicators	Baseline	2020 Target	Unit	Regional Benchmark	International Benchmark
Average student results in international TIMSS tests (eighth grade: Math and Science)	394 (Math), 436 (Science)	450 (Math), 480 (Science)	Number	452, 483	611, 590
Average student results in international TIMSS tests (fourth grade: Math and Science)	410 (Math), 429 (Science)	460 (Math), 470 (Science)	Number	469, 463	606, 583
Average student results in international reading tests (PISA)	Not Applicable	460	Number	475	542
Average student results in international reading tests (fourth grade – PIRLS)	430	460	Number	439	567

As shown in Table 1, this ambitious vision specifies for the first time a baseline performance, 2020 target performance, regional, and international benchmarks. In mathematics, for example, the targets by 2020 for both eighth and fourth grades in TIMSS are 450 and 460 scores respectively.

LITREATURE REVIEW

TEACHER COLLABORATION AND STUDENTS' OUTCOMES

During the last three decades, scholars have studied teacher collaboration through different lenses. In this section, the most relevant studies will be reviewed. Interestingly, teacher collaboration is found under different terms across studies. Terminologies that have been frequently used to study collaboration include collaborative learning communities (Cooper & Boyd, 1997; Haberman, 2004), communities of practice (Lima, 2001), professional learning communities (Dufour, 2004; Lomos, Hofman, & Bosker, 2011), teacher team (Hackmann, Petzko, Valentine, Clark, & Nori, 2002), critical friends groups (Curry, 2008; Hargreaves, 2001), and teacher collaboration (Goddard et al., 2007; Honingh & Hooge, 2014; Leonard & Leonard, 2003; Miller et al., 2010; Ronfeldt et al., 2015). Although these terms are sometimes used interchangeably (Vangrieken et al., 2015), the current study will focus on studies using the term teacher collaboration to examine its relationship to student outcomes.

Although teacher collaboration plays a critical role among key components of effective school culture (Little, 1982; Lortie, 1975), scholars have recently started examining its impact on student outcomes. Goddard et al. (2007) for instance, conducted a study to examine the link between collaboration and student achievement at 47 schools in one of the largest urban school districts in the mid western United States. Using the hierarchical linear modeling (HLM) statistical technique, the researchers found that teacher collaboration was a significant positive predictor of differences among schools in student achievement. Utilizing TIMSS 2003 dataset, Akiba et al. (2007) compared 46 countries regarding the opportunity gap to students' access to qualified mathematics teachers (teachers with full certification, a mathematics or mathematics

education major, and at least three years of teaching experience) and the student achievement gap. They found that the opportunity gap of qualified teachers was not significantly associated with achievement gap between the high and low socioeconomic status of students. To narrow the achievement gap they recommended that more attention to other effective instructional practices among teachers is needed.

This finding was later supported by another study that took place in one of Chicago's school districts (Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010). This study's aim was to examine the methods in which school practices and school and community conditions promote or inhibit students' performance in mathematics and reading. They found that schools that are effective in improving student learning tend to have, among several structures, a strong professional capacity. When the researchers deeply examined professional capacity in schools, they discovered that the individual qualifications of teachers were not nearly as important as the way teachers worked together. They concluded that schools with strong collaboration were more effective than schools with strong individuals but little collaboration (Bryk et al., 2010). In a follow-up study, Miller, Goddard, Goddard, Larsen, & Jacob (2010) tested the link between student achievement and a broader measure of collaboration, which included teachers' ratings on three different scales: frequency of collaboration on instruction, the extent to which teachers collaborate on instructional policy, and teachers' participation in formal collaboration structures. Utilizing the data from the first year of the School Leadership Improvement Study (SLIS) that included 96 elementary schools located in the northern regions of a Midwestern state serving students in rural areas, and applying structural equation models, the authors reported a direct effect of teacher collaboration on student achievement and an indirect effect of principals' instructional leadership on student achievement mediated by collaboration.

Horizontal collaboration (Hargreaves, 2001) that takes place among teachers in the same grade or departments is another research line in teacher collaboration literature. Lomos et al. (2011) examined empirically the relationship between secondary math departments in Dutch schools and student achievement. The study's results yielded a positive effect size of .20 showing those departments that focus on reflective dialogue, collaborative activity, and shared vision are associated with successful schools and higher student achievement.

In several other studies, the relationship between teacher collaboration and student achievement outcomes was found to be indirect. Drawing from a sample of 325 teachers and 2,270 students in an initiative that took place in a suburban school district in Connecticut, Zito (2011) examined the relationship between teacher collaboration and two dependent variables: changes in teachers' instructional practice and student achievement outcomes. He concluded that this relationship was indirect and mediated by strong administrative support.

Recently, Goddard et al. (2015) supported Zito's finding. In a sample of 93 elementary schools serving students in rural, high-poverty areas located in the northern regions of a Midwestern state, researchers were able to detect an indirect positive and significant relationship between both instructional leadership and formal teacher collaboration on one hand and student achievement on the other hand through their effects on collective efficacy beliefs. In another study that utilized the Early Childhood Longitudinal Survey (ECLS-K), researchers investigated two vital school culture factors, teacher professional communities and collaboration (Banerjee, Stearns, Moller, and Mickelson, 2016). They found that components of school culture, namely teacher professional community and teacher collaboration were significantly moderate the relationship between teacher job satisfaction and student achievement growth in both reading and mathematics.

More recently, in a large-scale study conducted in Miami-Dade County public schools over two years, Ronfeldt, Farmer, McQueen, and Grissom (2015), surveyed 9000 teachers and 336 schools to examine the types and the quality of collaborations that exist in instructional teams and whether these collaborations predict student achievement. By running a series of OLS regression models, researchers provide evidence that the average quality of faculty collaboration is associated with school achievement in both mathematics and reading value-added scores. Further, they found that female teachers report better quality collaboration in general than males. Noteworthy, researchers suggested for a future research a more probe of the psychometric properties involved in measuring collaboration quality.

Qualitatively, a group of researchers in Belgium examined teachers' storytelling, helping, sharing and joint work with regard to teachers' use of pupil learning outcomes, and what teachers learn from these learning activities (Van Gasse, Vanlommel, Vanhoof, & Van Petegem; 2016). Within the six teams they investigated, teachers primarily engaged in learning activities that incorporate no or little interdependency. Sharing and joint work with regard to teachers' use of pupil learning outcomes, and learning activities that imply a higher degree of interdependency were found to be rare.

In the same vein, Forte & Flores (2014) conducted a study in Northern Portugal sought more understanding of how teachers relate opportunities for professional development to their collaborative work in the workplace, and how they describe their collaborative experiences in the workplace. Applying mixed methods, they found that organizational factors such as working conditions, lack of training in collaboration, time, and motivation are the most limitations in relation to collaborative work.

In a recent and comprehensive systematic review of teacher collaboration, a group of researchers documented several critical points. Of relevance is that their caution about possible negative consequences of collaboration in individual and group levels (Vangrieken, Dochy, Raes, & Kyndt, 2015). In the individual level, teachers may experience competitiveness, tensions that can raise conflicts, a loss of autonomy, and increased workload. Possible negative consequences on the group level include for example groupthink, increased incompatibilities, more focusing in practical matters, and balkanization that is a new form of isolation on the group level. Moreover, they described four different interaction processes that show a hierarchy in levels of depth of teacher collaboration. First of these processes is individualism where the teachers' focus is on their responsibility and autonomy. Second is the perceived coordination through which teachers coordinate their responsibilities and tasks without further discussion of the substance of teaching. Third is the cooperation stage in which teachers establish a common ground for joint enterprise through focusing on the content and process of classroom activity. Finally sharing by which teachers begin to build, share, and clarify their pedagogical motives that direct the way they teach and learn.

Of particular relevance are studies that analyzed individual items of TIMSS teacher collaboration measure. One example of these studies is Blömeke, Vegar and Suhl in Nilsen & Gustafsson (2016) study that used TIMSS 2011 to investigate the relationship between teacher quality, instructional quality, and student achievement in mathematics for fourth grade across 47 countries. They used item parceling to construct their instructional quality measure. Among several items of this measure were CIT and IGS, two of the current study measures. Applying SEM, researchers concluded that across all countries, mathematics achievement of students at

grade four was not predicted by instructional quality, and within countries the predictor had a significant relation to achievement in only three countries.

In a follow-up study, Nilsen & Gustafsson (2016) investigated relations between within-country change in emphasis on academic success (SEAS), one of the current study's mediator (ACS), teacher quality, and change in mathematics achievement for 38 educational systems participating in the TIMSS grade eight assessments in 2007 and 2011. Among several findings, the researchers found that SEAS did not satisfy ideals of unidimensionality. More importantly, they found that SEAS did not predict student mathematics achievement scores. When they examined its items individually, only items reflecting parental support for student achievement and students' desire to perform well were significantly related to student achievement.

Sturman & Lin, (2011) Utilizing TIMSS 2007 to compare between the eight grade mathematics performance in England, Chinese Taipei, Republic of Korea, Singapore, Hong Kong, and Japan. Among different variables they investigated were two items of the teacher collaboration measure; discussed how to teach a particular concept, and prepared teaching materials together. The researchers detected only in one country, England, a small effect of these two teacher interactions on student mathematics achievement scores. They concluded that the impact of these interactions depends on different countries' circumstances.

Another recent example is Reeves, Pun, & Chung (2017) study that utilized TIMSS 2011 data to determine whether five unique indicators of collaboration predicted student achievement and teacher's job satisfaction in Japan and the United States. The study found that while collaboration during planning was the only indicator that was a significant predictor of student achievement in the United States, surprisingly, none of the five indicators of collaboration significantly predicted student achievement in Japan.

By mapping the terrain of current teacher collaboration literature, it is obvious that this concept has been differently operationalized across several studies. Yet, there are several gaps that call for further investigations before substantive conclusions can be made about the robustness of teacher collaboration and how it can be a worthwhile investment for both policymakers and practitioners. For example, more evidence gathered in different contexts is needed to further establish the link between teacher collaboration and not only student achievement, but also other student outcomes. Furthermore, most of studies conducted examined the formal collaboration that structured by schools or school districts, which differs from the informal collaboration (Darling-Hammond & Sykes, 1999; Goddard et al., 2015) that is the focus of the present study. Moreover, as shown in the aforementioned studies, although the well-known database of TIMSS that includes more than 50 measures including a measure of teacher collaboration, studies that utilizes these dataset are rare. Finally and with little exceptions, the majority of studies that examined teacher collaboration and its relation to students' outcomes were conducted in Western countries, and the Far East of Asian countries, which left the MENA countries behind in this research area.

CHAPTER 3

METHOD

3. 1 Overview

The purpose of this study was to empirically investigate the association between teacher collaboration and student outcomes. To achieve this goal, this cross-sectional study derived its data from TIMSS 2011. Since its inception in 1995, this International large-scale assessment (ILSA) data has gotten the attention of researchers, policy makers, and practitioners as well. According to The International Association for the Evaluation of Educational Achievement (IEA), an independent institution that partly sponsors TIMSS, the core objective of ILSAs is to help educators across the world learn what pedagogical approaches are more effective in the classroom (Martin & Mullis, 2008). Recently, among other factors (database, organizations, and individuals), TIMSS had the second-highest overall influence index value affecting the educational system in the United States (Swanson & Barlage, 2006). Student engagement, their confidence in learning mathematics, and their mathematics achievement scores were the focus of this study. Out of other variables, student engagement and self-confidence in learning are key factors of academic success (Kloosterman, 1988; Cai & Liem, 2017). Mathematics subject was chosen because of its significance as fundamental skills of student learning, particularly in early grades. Further, there is a national focus on mathematics within the educational initiatives in the 2030 Saudi Vision, the most recent and comprehensive reform movement in Saudi Arabia. In addition, it has been suggested that academic success in early grades is found to be a strong predictor of success in later grades (Dee, 2004; Tach & Farkas, 2006). Therefore, elementary students (fourth graders) were the focus of the current study.

3. 2 Target Population

The target population of this study includes all students enrolled in the fourth year of formal schooling in Saudi Arabia in 2011. As many other countries, this fourth year of formal schooling as defined by UNESCO International Standard Classification of Education (ISCED) (Rutkowski, Gonzalez, Joncas, & Von Davier, 2010) meets the fourth grade in Saudi educational system. Locally, this grade is named elementary Year 4. Students' average age in this grade was 10.0. In TIMSS 2011, the student population was 401,006 studying in 11,394 schools. While the aim of TIMSS was a comprehensive coverage of the target population, it was difficult to attain. In some cases and for different reasons, there was exclusion. However, the overall number of exclusion (including schools and students) must be less than 5% of the national target population in each country (Rutkowski et al., 2010). The exclusion in Saudi target population in TIMSS 2011 within school, student, and overall was 1.2%, 0.1%, and 1.2%, respectively.

3. 3 Sample

TIMSS fundamentally is a study of student achievement; therefore, the precision estimation of student achievement through precise sample procedures is of primary importance. To provide a representative sample from the target population in Saudi Arabia in TIMSS 2011, a two-stage stratified cluster sample design was utilized (Martin & Mullis, 2013). In this sampling technique, schools were initially sampled with the probability proportional-to-size (PPS), by which the larger the size of a school, the greater the chance of being selected in the sample.

There were two stratification procedures to sample schools in Saudi Arabia: explicit and implicit stratification. In explicit stratification, a separate list of schools (sampling frame) was constructed based on school gender (Boys, Girls) and school type (Public, Private). Within the implicit stratification, Saudi schools were sampled geographically in five regions (Center, East,

South, West, North), school type (Public, Private), and school gender (Boys, Girls). According to Joncas and Foy (2013), combining the implicit strata and explicit sampling was an effective way to obtain a proportional sample and leads to improve the reliability of the achievement estimates, given that the implicit stratification variables were correlated with student achievement (Martin & Mullis, 2013).

Practically, schools from each explicit stratum in the sampling frame were sorted by implicit stratification variables and by their measure of size (MOS). Eventually, 168 schools were sampled from the whole Saudi school target population. In the second stage, one or more intact classes from the fourth grade in the sampled schools were randomly sampled. With consideration of class size, this stage yielded a representative sample of 3773 of fourth-grade students taught by 168 mathematics teachers. This study capitalized on the nationally representative sample of students and schools from TIMSS 2011, which allowed for generalizations of the results of to the national level.

3. 4 Data Collection

Working in ILSAs such as TIMSS (2011), it is crucial to understand the characteristics of this ambitious international assessment, including its complex procedures for sampling, measuring students' achievement, and analyzing the data. These details are described in the TIMSS-2011 Methodology and Procedures chapter within the user guide (Martin & Mullis, 2013). In this section, instruments used to collect the data will be described in more details.

3. 5 Study Variables and Instruments

TIMSS (2011) administered two main parts of data. The first part is the achievement test scores (math, science), and the second part is the questionnaires of students, parents, teachers, principals, and curriculum designers. This study utilized student achievement scores in

mathematics, student engagement, and student confidence in learning as outcome variables. The main latent factor of this study was collaboration between teachers to improve teaching (CIT). Furthermore, the study examined two other latent factors that have previously been found in the literature to be mediators of the impact of the CIT on student outcomes. These two measures were teacher's Emphasis on Academic Success (ACS) and Instruction to Engage Students in Learning (IGS). In the following section, the study variables will be discussed.

3. 5. 1 Endogenous variables. The outcome variables in this study were student achievement scores in mathematics, their engagement, and their confidence in learning. In the following section, these outcomes and measures will be illustrated.

3. 5. 1. 1 Achievement scores. The first outcome is mathematics achievement scores for fourth grade students. These scores are based on the students' answers of multiple choices and construct questions. These questions covered content and cognitive topics. In mathematics content, for example, students were asked over numbers, geometric shapes and measures, and data analysis. In the cognitive domain, questions covered knowing, applying, and reasoning.

Although TIMSS 2011 had a large amount of questions to administer to the students, it was neither easy nor practical to ask students all of these questions. To allow for a large amount of data to be collected without burdening individual students from answering all of the possible questions, a matrix sampling approach was used (Martin & Mullis, 2013). In this approach, the pool of questions in mathematics was broken up into 14 blocks, which allowed for 14 test booklets in total. Each test booklet consisted of two blocks of mathematics and two blocks of science. This allowed for each student to only answer a portion of the whole questions in each subject.

For more accurate estimation of student achievement, TIMSS 2011 used Item Response Theory (IRT). In this technique, students' abilities to correctly respond to the questions were conditioned to the questions they answered and their background information (Martin & Mullis, 2013). Due to the different types of question responses, three different IRT models were used to calculate student achievement scores for the TIMSS 2011. It is important to note, however, that the scores reported by the TIMSS through this technique cannot be considered individual student test scores. Instead, they should be considered imputed scores (plausible values). This kind of score requires analysis in a certain way that will be explained later in the data analysis section of this study.

The second part of data instruments was the questionnaires developed for the TIMSS 2011. They include a: (a) student questionnaire, (b) teacher questionnaire, (c) school questionnaire, (d) parent questionnaire, and (e) curriculum questionnaire. This study analyzed the data collected from student, teacher, and school principals. In their questionnaire, students were asked about the home and school environment encountered and their learning experiences in classrooms. The teacher questionnaire was given to the teachers of the sampled students. They were asked about their background, preparation for teaching, their school experiences, professional development activities, their collaboration, and so on. The third questionnaire was given to school principals and asked for demographic information, school contexts, climate, and available resources for instruction (Martin & Mullis, 2013). In the following section, the measures that were derived from these questionnaires will be discussed.

3. 5. 1. 2 Student engagement measure (ENG). This construct was measured by asking students in a four-scale response (agree a lot, agree little, disagree little, and disagree a lot) the following question:

How much do you agree with these statements about your mathematics lessons?

1. I know what my teacher expects me to do.
2. I think of things not related to the lesson.
3. My teacher is easy to understand.
4. I am interested in what my teacher says.
5. My teacher gives me interesting things to do.

3. 5. 1. 3 Student confidences in learning (CON). This construct was measured by asking students in a four-scale response (agree a lot, agree little, disagree little, and disagree a lot) the following question:

How much do you agree with these statements about mathematics?

1. I usually do well in mathematics.
2. Mathematics is harder for me than for many of my classmates.
3. I am just not good at mathematics.
4. I learn things quickly in mathematics.
5. I am good at working out difficult mathematics problems.
6. My teacher tells me I am good at mathematics.
7. Mathematics is harder for me than any other subject.

3. 5. 2 Exogenous variables.

3. 5. 2. 1 Collaborate to improve teaching (CIT). CIT was the main predictor in this study. Teachers in TIMSS 2011 were asked about other professional learning opportunities beyond the regular professional development programs which might enhance their teaching quality (Martin, Mullis, Gregory, Martin, & Hoyle, 2003). Specifically, they were asked the following question:

How often do you have the following types of interactions with other teachers?

1. Discuss how to teach a particular topic.
2. Collaborate in planning and preparing instructional materials.
3. Share what I have learned about my teaching experiences.
4. Visit another classroom to learn more about teaching.
5. Work together to try out new ideas.

Teachers were asked to respond to these statements using a four-point scale (never or almost never, 2-3 times per month, 1-3 times per week, daily or almost daily).

3. 5. 3 Mediator variables. Consulting the literature and exploring Almanacs of TIMSS 2011, two mediators were included to the study model. These two mediators are ACS and IGS.

3. 5. 3. 1 School emphasis on academic success (ACS). This construct was measured by asking teachers in a five-scale response (very high, high, medium, low, very low) the following question:

How would you characterize each of the following within your school?

1. Teachers' understanding of the school's curricular goals.
2. Teachers' degree of success in implementing the school's curriculum.
3. Teachers' expectations for student achievement.
4. Parental support for student achievement.
5. Students' desire to do well in school

3. 5. 3. 2 Instruction to engage students in learning (IGS). This construct was measured by asking teachers in a four-point scale response (every or almost every lesson, about half of the lessons, some of the lessons, and never) the following question:

How often do you do the following in teaching this class?

1. Summarize what students should have learned from the lesson.
2. Relate the lesson to students' daily lives.
3. Use questioning to elicit reasons and explanations.
4. Encourage all students to improve their performance.
5. Praise students for good effort.
6. Bring interesting materials to class.

3. 5. 4 Background variables. After looking at the literature and exploring the Almanacs of Saudi TIMSS 2011 data, we also included three control variables to prevent potential biases in our analyses. These variables are student's SES (Baker et al., 2002; Coleman, 1966), leadership activities that have been found to be related to several school and teacher characteristics (Gustafsson, Nilsen, & Hansen, 2015), and previous learning, which has been considered influential in student achievement (Akiba, LeTendre, & Scribner, 2007; UNICEF, 2007). For the latter, attended pre-school was planned to use as a proxy for the prior achievement. However, the missing rate of this variable in the sample was unfortunately high (about 50%); therefore, it was eliminated from the study. In the following section, the other two control variables will be explained.

3. 5. 4. 1 School leadership activities. In TIMSS 2011, this construct was measured by asking school principals in a three-scale response (no time, some time, a lot of time) the following question:

During the past year, approximately how much time have you spent on the following school leadership activities in your role as a school principal?

1. Promoting the school's educational vision or goals.
2. Developing the school's curricular and educational goals.

3. Monitoring teachers' implementation of the school's educational goals in their teaching.
4. Monitoring students' learning progress to ensure that the school's educational goals are reached.
5. Keeping an orderly atmosphere in the school.
6. Ensuring that there are clear rules for student behavior.
7. Addressing disruptive student behavior.
8. Creating a climate of trust among teachers.
9. Initiating a discussion to help teachers who have problems in the classroom.
10. Advising teachers who have questions or problems with their teaching.
11. Visiting other schools or attending educational conferences for new ideas.
12. Initiating educational projects or improvements.
13. Participating in professional development activities specifically for school principals.

3. 5. 4. 2 Student's socioeconomic status (SES). This construct was measured by student responses to a question concerning whether they possess variety of items at home. Students were asked the following binary (yes, no) question:

Do you have any of these things at your home?

1. Computer
2. Study desk/table for your use
3. Books of your very own (do not count your schoolbooks)
4. Your own room
5. Internet connection
6. Motorcycle

7. Bicycle
8. Television
9. DVD
10. Play Station
11. Mobile phone

While the first five items of this measure were prepared by TIMSS 2011, the last six items were left to be specified by participated countries. In this cycle of TIMSS, therefore, items in this measure differed from country to country. Saudi Arabia added items 6 through 11 in this scale to measure their students' SES. Most items that included in this measure have widely been used to assess students SES in different studies (Akiba et al., 2007; Hauser, 1994).

3. 6 Analysis Software

The data was obtained form The International Association for the Evaluation of Educational Achievement (IEA), an independent international cooperative of national research institution that conducts TIMSS (<http://www.iea.nl/>). The IEA's International Database (IDB) Analyzer was used (IEA, 2008). While the IDB analyzer was developed to analyze the IEA's data, in the current study it was used to facilitate liking students to their teachers and schools. Next, the statistical package for the social science (SPSS) Version 25.0 (2015) was used for preparation purposes (e.g., data recoding and data error-checking). To analyze the data, Multi Confirmatory Factor Analysis (ML-CFA) and Multi Structural Equation Modeling (MSEM) procedures were carried out using Mplus Version 7.0 because of its specialization in modeling structural equation with latent variables and its capability to handle complex samples (Muthén & Muthén, 1995-2015). In addition, because of the model size that is based on the latent and observable variables, a high computing procedure through the cluster in The Center for Research

Computing (CRC) available through (<https://researchcomputing.ku.edu/center-research-computing>) was used to facilitate models' convergence.

3. 7 Analysis Methods

Based on this study's conceptual framework, it has been hypothesized that teacher collaboration was associated directly and indirectly to student outcomes. Unlike past research that utilized manifest variables based on aggregated scale scores, the constructs of the current study were treated as latent variables. Thus, a two-level structural equation model (SEM) was utilized to control for measurement error. On the bases of previous findings described in the literature and the conceptual framework of this study, the researcher hypothesized the following relationships as depicted in the structural model presented in Fig. 1.

Hypothesis 1. Teacher collaboration (CIT) was positively and significantly associated with differences among schools in students' fourth-grade level of mathematics achievement scores, engagement, and confidence in learning.

Hypothesis 2. This effect of teacher collaboration was mediated through teachers' emphasis on Academic Success (ACS) and instruction to engage students in learning (IGS). Because no more than two intact classes were sampled from each school in TIMSS 2011 Saudi data, the differences between classes were indistinguishable from the differences between schools. As a result, schools were standing for the schools/classes level beside the student level, which produced a two-level analysis of the data. Due to TIMSS test design that was mentioned in chapter three, (Martin & Mullis, 2012), five plausible values were randomly drawn from a proficiency distribution based on both achievement scores on test items and on student background variables. Therefore, the analysis of the current study's data was run five times using a different plausible value

each time. Then, the average of the five estimates was used (Rutkowski, Gonzales, Joncas, & von Davier, 2010; Wang, 2001). This procedure is facilitated in Mplus using a multiple imputation model (Muthén & Muthén, 1998–2015). To facilitate numeric stability, the plausible values were divided by 100.

3. 7. 1 Measurement model. *Given the measurement limitations of TIMSS scales (Hooper, Arora, Martin, & Mullis, 2013; Marsh et al., 2013) and the need for high-quality measurement particularly for SEM analysis (Martin & Mullis, 2011; Shadish, Cook, & Campbell, 2002) the reliability and dimensionality of the seven constructs of the current study were assessed through confirmatory factor analysis using the robust maximum likelihood estimation (MLR) for continuous items, and a limited-information weighted least squares (WLSMV) for categorical items through Mplus v. 7.4 (Muthén & Muthén, 1998-2015). Continuous item models were identified by fixing each latent factor variance to one and means to zero such that all item intercepts, item factor loadings, and item residual variances were then estimated. Further, categorical (the threshold model) item models were identified by fixing each latent factor variance to one and means to zero such that all thresholds and loadings were then estimated. Furthermore, across these models, unique variances were uncorrelated unless specified. In addition, the models were also constrained so there were no cross-loading indicators. Covariance among the latent factors in each level were included. Prior to running the analysis, several items in different scales were reverse-coded such that higher values indicated greater levels of the latent trait for all items. Several model fit indices were utilized to assess models' fit. These indices include the likelihood ratio test statistic (a chi-square χ^2) statistic that represents the discrepancy between the unrestricted sample covariance matrix (S) and the restricted covariance matrix $\Sigma(\theta)$, its degrees of freedom, its*

p-value (in which non-significance is desirable for good fit), the Root Mean-Square Error of Approximation (RMSEA) point estimate, and 90% confidence interval (in which values lower than .06 are desirable for good fit) (Schreiber, Nora, Stage, Barlow, & King, 2006). In addition, the Standardized Root Mean Square Residual (SRMR) that indicates the amount of deviation between the elements of the observed covariance matrix and the model-implied matrix will be also used to assess the measures fit (Schreiber et al., 2006); this measure should be .08 or lower. The Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI) are other two indices used to assess the model goodness of fit. Both indicators should be $> .95$ (Schreiber et al., 2006).

3. 7. 2 Structural model. Structural equation modeling (SEM) is considered as the most appropriate method to study causal mechanisms using latent variables (Muijs, 2012). It has several advantages over other statistical procedures. First, SEM offers an enhanced understanding of the mechanism of complex relationships that exist among theoretical constructs. It also provides a comprehensive method for specifying and empirically testing the plausibility of such complex and theoretical models (Kelloway, 1998). Other advantage is that it allows for capturing the direct and indirect effects with multiple exogenous and endogenous factors simultaneous (Crockett, 2012). The difference between exogenous and endogenous factors is that while the later represents model factors that are influenced only by the former, the former (exogenous) represents factors in the model that are influenced by different other factors that are not part of the model. In the current study's level one (students), engagement in math (ENGM), confidence in learning math (CONM), and math achievement scores (PVM) were considered endogenous factors while student socioeconomic status (SES) was considered an exogenous factor. In level two (schools), the random intercepts of ENGMB, CONMB, and PVM are endogenous. Instructional leadership (LDINS), leadership activities that emphasis student

discipline (LDDIS), and school SESB are exogenous factors. Finally, teacher collaboration (CITB), teacher emphasis on academic success (ACSB), and teacher instruction to engage student (IGSB) are simultaneously exogenous and endogenous variables. Another final advantage of SEM is that this statistical procedure does not just assume that no measurement error exists for the exogenous variables, but it also tests this assumption empirically not only in a single level, but also in a multilevel framework such the current study's nested data (Brown, 2006; Crockett, 2012). Ignoring potential measurement errors can lead to a biased estimation of the standard errors. For these advantages and in consideration of the study's conceptual framework and its data structure, SEM was deemed a suitable statistical technique using Mplus v. 7.4 (Muthén & Muthén, 1998-2015). To account for the two-stage stratified cluster sample design of TIMSS 2011 (Martin & Mullis, 2012), a two-level structural equation model (MSEM) with latent variables (Hox et al., 2010) was estimated. The structural relationships in both levels, students and schools, were described. However, the main focus of the study was on the teacher/school level rather than on the student level since the main predictor, teacher collaboration, resides in the school level. Specifically, the study applied the multilevel structural equation modeling (MSEM) with a random intercept to investigate the effects of teacher collaboration on student engagement (ENG), student confidence in learning (CON), and student mathematics achievement scores (PVM). Further, the study examined if the effect of teacher collaboration on ENG, CON, and PVM was mediated by teacher emphases on academic success (ACS) and their instruction to engage students (IGS). These indirect effects were estimated using the MODEL CONSTRAINT command, in which the NEW option is introduced to new parameters that are not part of the MODEL command.

The study estimated random intercept variances for the latent constructs. Specifically,

this model allowed the intercepts of the level one (students) factors to be predicted by level two (schools/ teachers) factors. In a two-level model with missing data and unbalanced group sizes (as is the current study's data), it is essential to apply an estimator that considers these conditions. Because the current study has a combination of continuous, binary, categorical (ordinal) variables, a limited-information weighted least squares (WLSMV) applied (Asparouhov & Muthén, 2007). Based on TIMSS data structure and the clustering procedures, a correction to the standard errors and Chi square test of model fit applied to consider the nested structure of the data using the Mplus TYPE=TWOLEVEL and CLUSTER commands (Muthén and Muthén, 1998-2015). Moreover, for technical reasons, TIMSS applies IRT methodology to obtain students achievement scores. To do so, it imputes five scores, also referred as plausible values, for each student. Because of the uncertainty that might exist in this imputation, the researcher run the analysis of the data once for each plausible value, for a total of five times. The average of these five sets of results has been used as the best estimate for the analysis of student achievement (Foy & Olson, 2009; Rutkowski, Gonzalez, Joncas, & von Davier, 2010). This procedure carried out using TYPE=IMPUTATION available through Mplus (Muthén and Muthén, 1998-2015). Because of this approach, however, the data collected can make inferences about the distribution of student achievement in the country rather than that of the individual student (Foy, Galia, & Li, 2008). Furthermore, missing data is a serious concern in SEM. Therefore; a detailed description of this issue was reported based on the pattern of missing data and how it was handled in the study.

CHAPTER 4

STUDY FINDINGS

The purpose of this study was to test the sufficiency of the conceptual model of the effects of teacher collaboration (CIT) on student outcomes using Structural Equation Modeling (SEM). Specifically, the present study sought to test the following hypotheses:

Hypothesis 1. Teacher collaboration (CIT) is positively and significantly associated with differences among schools in students' fourth-grade level of achievement in mathematics, engagement, and confidence in learning.

Hypothesis 2. Teacher collaboration is positively and significantly associated with differences among schools in students' fourth-grade level of achievement in mathematics, engagement, and confidence in learning through teachers' emphasis on Academic Success (ACS) and instruction to engage students in learning (IGS).

In order to investigate the sufficiency of the theoretical model and to calculate the total estimated effect of teacher collaboration (CIT) in student outcomes, the study utilized multilevel structural equation modeling (MSEM) to test the hypothesized relationships between these variables. Prior to conducting the analysis, missing data pattern was examined using PATTERNS and FREQUENCIES commands in Mplus output option. These two output commands showed that the attending a pre-schooling program variable that was planned to be included in the model as proxy for a previous achievement, which should be controlling for, has more than 50% missing values; therefore, it was eliminated from the study (Goddard et al., 2015). Further, because the main predictor of the current study was in the second level (school/ teacher), and to apply the random intercept model, Mplus requires the variation to be only between, not within

clusters; therefore, in schools that were sampled two classes, the smaller one was deleted. The amount of data deleted by this procedure was negligibly small. Otherwise, no patterns of missing data were found. Based on that, missing at random (MAR) was assumed. In addition, several items in student level (ENG, CON) were reverse-coded prior to the analysis such that higher values then indicated greater levels of the latent trait for all items. After data cleaning, 3773 students nested in 168 schools with an average cluster size of 22.45 were included in the analysis. TIMSS 2011 basic sampling design is a two-stage cluster design consisting of sampling of schools and intact classrooms from the target grade in the school. Participants for the current study were 3,773 students (out of 401,006) fourth-grade students from 168 (out of 11,393) elementary schools in Saudi Arabia. The students' age average in this grade was 10.0. The participants were mostly distributed equally in gender (51.6% female).

4. 1. Measurement Model

To take advantage of applying SEM, a high-quality measure should be first secured. Given the limitations of TIMSS scales (Hooper, Arora, Martin, & Mullis, 2013; Marsh et al., 2013) the reliability and dimensionality of the seven constructs of the current study were assessed through confirmatory factor analysis. Specifically, the study investigated the psychometric properties of seven latent factors in two steps. First, a MCFA of the level one (student level) latent traits, which are student engagement in mathematics (ENG), confidence in learning mathematics (CON), and socioeconomic status (SES), was conducted in a sample of 3,773 fourth-grade students. Prior to proceed to MCFA, the need of a multilevel procedure was examined by investigating the variability between and within schools on each item of these three constructs that was inspected by computing the intraclass correlations (ICCs). The ICCs for the observed variables provide a measure of the amount of variability between schools and the

degree of non-independence or clustering of the data within schools. Using a random intercept model, the ICC for an item represents the variation between schools in the intercepts (means) of the item divided by the total variation (sum of the variation between schools in the intercepts and the variation within schools). ICCs can range from 0 to 1.0, with larger values indicating greater clustering effects within schools. There is no clear-cut point for how large the ICC needs to be to warrant multilevel analyses. While some researchers consider an ICC of .05 an indicator of multilevel (Huang & Cornell, 2016; Musca et al., 2011), other researchers consider an ICC of 0.1 or greater as enough evidence of the multilevel structure to be modeled (Dedrick & Greenbaum, 2011; Dyer, Hanges, & Hall, 2005; Little, 2013; Muthén, Kao, Burstein, Muthen, Kao & Burstein, 1991). Utilizing Mplus 7.4 version (Muthén & Muthén, 1998-2015), Table 2 displays a sample of the highest and lowest ICCs for ENG, CON, and SES.

Table 2

Interclass Correlation for the Highest and Lowest Level One Construct Items

ITEM	ICC
SES11	0.37
SES7	0.15
ENG2	0.23
ENG4	0.20
CON2	0.18
CON6	0.07

The ICC values show variability ranging from 37% in SES to 7% in the confidence construct. These values indicate that these constructs should be modeled in a multilevel CFA. The second step of the measurement model was assessing the latent construct measures in level two (schools). These measures are teacher collaboration (CIT), teachers' emphasis on academic success (ACS), teachers' instruction to engage students (IGS), instructional leadership activities

(LDINS), and leadership activities that emphasis student discipline (LDDIS). Therefore, several CFAs were conducted to assess the reliability and dimensionality of these seven constructs.

Using Mplus 7.2 (Muthén & Muthén, 1998-2015), two estimators were applied. First, WLSMV estimator was used to estimate categorical item of ENG and CON. Second, MLR estimator was used to estimate CIT, ACS, IGS, SES, IDINS, and IDDIS. Model fit statistics reported in Table 3 include the obtained model χ^2 , its degrees of freedom, and its p-value (in which non-significance is desirable for good fit), CFI, or Comparative Fit Index (in which values higher than .95 are desirable for good fit), and the RMSEA, or Root Mean Square Error of Approximation, point estimate and 90% confidence interval (in which values lower than .06 are desirable for good fit).

Table 3

Measurement Models

Measurement Models	Items	χ^2	CFI	TLI	RMSEA	SRMR W/B*
ENG	5/5	(df=10)= 64.631, P= 0.001	0.96	0.92	0.04	0.02/ 0.06
CON	7/6	(df=16)= 54.061, P= 0.001	0.98	0.97	0.03	0.02/ 0.04
SES	11/10	(df=66)= 264.27, P= 0.001	0.93	0.91	0.03	0.03/ 0.13
2-F LD	13/10	(df=32)= 48.21, P= 0.03	0.97	0.96	0.01	0.05
CIT	5/5	(df=5)= 1.298, P= 0.94	1.00	1.04	0.00	0.01
ACS	5/5	(df=4)= 12.24, P= 0.015	0.96	0.90	0.02	0.04
IGS	6/5	(df=4)= 4.033, P= 0.401	0.99	0.99	0.001	0.03

*SRMR W/B means within and between.

As shown in Table 3, except for ENG and CIT, all latent factors did not adequately describe the pattern of relationship across their items as initially hypothesized. Sources of local misfit were identified using the normalized residual covariance matrix, available via the RESIDUAL output option in Mplus, in which individual values were calculated as follows: (observed covariance – expected covariance) / SD (observed covariance). Relatively large positive residual covariances were observed among items ACSM1 and ACSM2 in ACS, IGSM4 and IGSM5 in IGS, SES1 with SES5 and SES4 with SES8 in SES, and item CON2 with CON7,

indicating that, within a given factor, these items were more related than was predicted. By reviewing these items, content overlapping or method effects (Abu-Hilal et al., 2014, 2013; Hooper et al., 2013) were detected; thus, these items could be covered.

In checking these items translation to the Arabic version, item CON3 and IGS3 were poorly translated. In specific, items CON3 had different meanings in the Arabic version. This item was “I am just not good at mathematics.” It was translated, however, to mean, “I am not good at mathematics only.” It is clear that replacing “just” with “only” and shifting it to another place in the sentence dramatically changed the meaning of the statement. Likewise, item IGS3 asked teachers to which extent they “use questioning to elicit reasons and explanations.” This item was translated to mean, “using questions to check their students’ understanding.” There is a difference between “questioning” and “asking questions.” While the latter can put people in a defensive situation, the first is more likely to make them relax and explain their rationale for a course of action or thought (Moyer & Milewicz, 2002). As a result, these two items were deleted from the two measures. Further, item SES6 from the SES measure was deleted for poor loadings.

The normalized residual covariance matrix indicated that there would be a need for separating leadership activities factor (LD) into two factors: instructional leadership (LDINS) with items LD1, LD2, LD3, LD4, and LD13 and leadership activities that emphasis student discipline (LDDIS) with items LD5, LD6, LD7, LD8, and LD9. LD8 with LD9 and LD1 with LD2 were covered. In addition, items LD10, LD11, and LD12 were deleted for poor loadings. The two-factor model fit solution of LD was acceptable, indicating that these leadership activities measure two separate but related constructs. Further examination of local fit via normalized residual covariances and modification indices yielded no interpretable remaining relationships among these seven factors.

Overall, standardized factor loadings are substantial and statistically significant for constructs at the individual student level and school level. Residual variance terms are close to zero. Thus, based on the current study's theoretical framework and these empirical results, the researcher would argue that these latent traits are unitary constructs, which suggests including them in the structural model.

Table 4

Correlation between Constructs in L1

	1	2	3	4
ENG				
CON	0.82*			
PVM	0.06	-0.24		
SES	-0.02*	0.26*	0.30*	

Table 5

Correlations between Constructs in L 2

	1	2	3	4	5	6	7	8	9
ENG									
CON	0.86*								
PVM	-0.34*	-0.40*							
SES	-0.42*	-0.39*	0.20*						
CIT	-0.12	0.05	0.10	-0.02					
ACS	0.29*	0.29*	-0.36*	-0.06	-0.21*				
IGS	0.60*	0.53*	-0.30*	-0.31*	-0.59*	0.60*			
LDS	-0.09	0.03	0.17	-0.01	-0.10	-0.04	0.03		
LDD	0.08	0.10	0.10	-0.15	-0.04	-0.03	0.05	0.62*	

* Sig. $p < .05$

Prior to run the MSEM model, correlations between constructs within each level were run and findings are reported in Tables 1 and 2. The zero-order correlations were not quite consistent with the hypothesized relations. Student engagement (ENG) and their confidence in learning (CON) were positively and significantly correlated ($r = 0.86$, $p < .05$); however, both were negatively and significantly correlated to student plausible values in mathematics (PVM) ($r = -0.34$ and -0.40 , $p < .05$). Furthermore, the main predictor of the study, teacher collaboration (CIT) was negatively and significantly correlated to ENG, but positively to CON and PVM ($r = -$

0.12, 0.05 and 0.10, $p < .05$). Unexpectedly, CIT was negatively and significantly correlated to the two mediators of the model, emphasis on academic success (ACS) and instruct to engage student (IGS) ($r = -0.21$ and -0.59 , $p < .05$). To test the study's hypotheses, two models were conducted: measurement model and structural model. The following two sections explain in detail these two models.

4. 2 Structural Model

The current study used multilevel structural equation modeling (MSEM) to test the relationships between teacher collaboration and their student outcomes in fourth grade. Specifically, it examined the direct relationship between teacher collaboration (CIT) and student engagement (ENG), student confidence in learning (CON), and student mathematic achievement scores (PVM). Also, as it hypothesized, the study tested whether these relationships were mediated by ACS and IGS.

The model displayed in Figure 2 depicts the interrelationships implied by the study's two research questions. The exogenous and endogenous variables are presented within ovals. Because these variables represent latent constructs measured by more than one item, the model considers them unobserved variables. Also, ENG, CON, and PVM were presented in the school level within ovals because they were random intercepts. As seen in Figure 1, CIT, ACS, and IGS

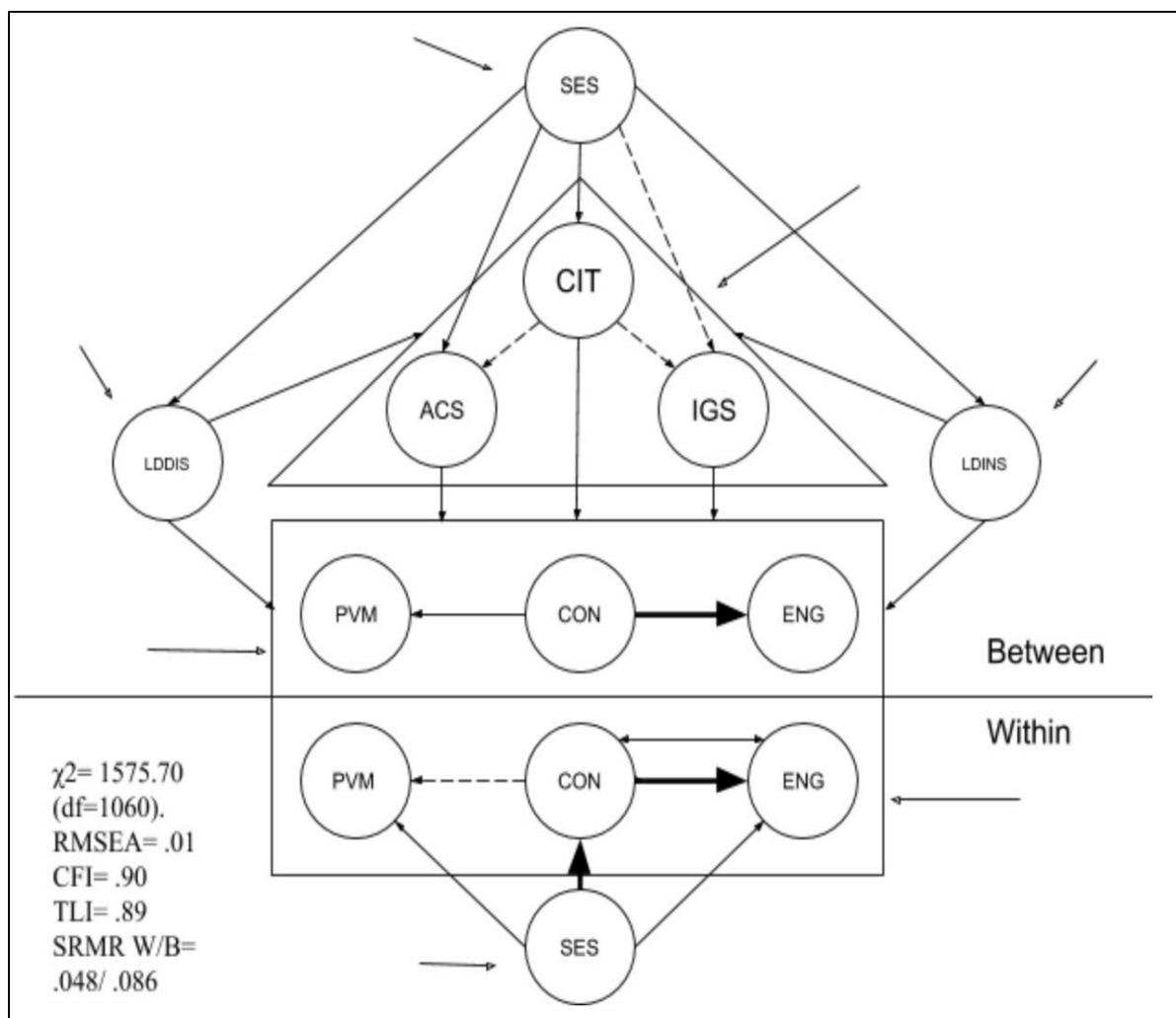


Figure 3. Multilevel SEM of direct and indirect relationship of Teacher Collaboration (CIT), Emphasis of Academic Success (ACS), Instruct to Engage Student (IGS), Student Engagement (ENG), Student Confidence in Learning (CON), and Student Math Achievement (PVM). For simplicity, Manifested items are not included.

- Positive Relationship
- ⇝ Negative Relationship
- Residuals

Were entered in the model as predictors of ENG, CON, and PVM. In this random intercept model, student outcome intercepts of ENG, CON, and PVM (level 1) could be predicted by CIT directly and ACS and IGS indirectly. Student socioeconomic status (SES) and school leadership activities (LD) were included in the model as covariates. The analysis of TIMSS plausible values of mathematics achievement scores was run once for each plausible value, for a total of five

times. The average of these five sets of data has been used as the best estimate for the analysis of student achievement. The robustness of the theoretical model was evaluated according to various fit indices. Because of the well-documented sensitivity of a chi-square test of fit (χ^2) to the large sample size (Brown, 2006; Byrne, 2012), the root mean-square error of approximation (RMSEA), the Tucker-Lewis Index (TLI), and Comparative Fit Index (CFI) were utilized to assess the model good fit (Hox, 2002; Hu & Bentler, 1999; Yu, 2002). The MSEM diagnostics showed the root mean square error of approximation (RMSEA = .02, SD= 0.0001), comparative fit index (CFI = .90, SD= 0.001), the Tucker Lewis index (TLI=0 .89, SD= 0.001), and the standardized root mean square residual (SRMR within =0 .040, SD= 0.0001) and SRMR between =0 .08, SD= 0.001). These fit indices indicated that the possible relationships in this model have been captured. Thus, this research proceeded with confidence to interpret the multilevel SEM parameter estimates for their substantive meaning in relation to the study hypotheses.

Table 4

Standardized Structural Relationships in Level 1

Latent Factors	β	SE
CON → ENG	0.82**	0.02
SES → CON	0.26**	0.03
CON → PVM	-0.24**	0.02
SES → ENG	-0.02	0.04
SES → PVM	0.03	0.03

* $p < .10$. ** $p < .001$ → Not Sig
 → Positively Sig → Negatively Sig

Table 5

Standardized Structural Direct Relationships in Level 2

Latent Factors	β	SE	Latent Factors	β	SE
CONB→ENGB	0.78**	0.21	LDINS→ACS	-0.06	0.18
CIT→ENGB	-0.11	0.29	LDDIS→ACS	-0.01	0.19
SESB→ENGB	-0.05	0.09	SESB→ACS	-0.06	0.11
LDINS→ENGB	-0.20	0.11	CIT→ACS	-0.22*	0.12
LDDIS→ENGB	0.11	0.12	LDINS→IGS	-0.04	0.21
ACS→ENGB	-0.04	0.20	LDDIS→IGS	0.01	0.23
IGS→ENGB	0.14	0.49	SESB→IGS	-0.32*	0.13
CIT→CONB	0.57	0.42	CIT→IGS	0.61**	0.14
SESB→CONB	-0.11	0.27	CONB→PVMB	-0.61	0.41
LDINS→CONB	0.04	0.21	ENGB→PVMB	0.24	0.35
LDDIS→CONB	0.03	0.21	ACS→PVMB	-0.35	0.24
ACS→CONB	-0.15	0.41	IGS→PVMB	0.27	0.59
IGS→CONB	0.92	0.70	CIT→PVMB	0.27	0.35
LDINS→CIT	-0.13	0.13	SESB→PVMB	0.10	0.13
LDDIS→CIT	0.04	0.15	LDINS→PVMB	0.30	0.16
SESB→CIT	-0.01	0.10	LDDIS→PVMB	-0.14	0.14
SESB→LDINS	-0.01	0.09			
SESB→LDDIS	-0.15	0.09			

* $p < .05$ ** $p < .001$ → Not Sig
→ Positively Sig → Negatively Sig

Table 6

Mediated Relationships Between CIT and ENG, CON, and PVM

Indirect Relationships	β	SE
CIT→ ⁺ ACS→ ⁺ ENG	0.01	0.02
CIT→ ⁺ ACS→ ⁺ CON	0.01	0.04
CIT→ ⁺ ACS→ ⁺ PVM	0.09	0.08
CIT→ ⁺ IGS→ ⁺ ENG	-0.03	0.09
CIT→ ⁺ IGS→ ⁺ CON	- 0.24	0.20
CIT→ ⁺ IGS→ ⁺ PVM	- 0.20	0.45

* $p < .05$ ** $p < .001$ →⁺ Not Sig
 → Positively Sig →⁻ Negatively Sig

The central focus of the study was whether fourth-grade students' engagement, confidence in learning, and achievement scores in mathematics are associated with their teachers' collaboration, and to which extent did this association is mediated by teachers' emphasis on academic success (ACS), and instruct to engage student in Learning (IGS). The structural model presented in Figure 2, table 4 and 5 provide an illustration of the standardized coefficient paths of structural models at both levels. As shown in Table 4, at the student level, SES positively predicted student confidence in learning (CON) $\beta = 0.26$ ($p < .001$). Surprisingly, the model did not reveal any association between SES and mathematics achievement scores. Moreover, while student confidence in learning (CON) positively predicted student engagement (ENG) $\beta = 0.82$ ($p < .001$), accounting for 67% of the variance in ENG, it negatively predicted mathematics achievement scores (PVM) $\beta = -0.24$ ($p < .001$). In the school level, where the focus of this study is, table 5 unexpectedly shows that the model did not find any significant association between teacher collaboration and ENG, CON, and PVM (Hypothesis 1). More surprisingly, the indirect path analysis did not detect any mediating effects between CIT and ENG, CON, and PVM through teacher emphasis on academic success (ACS) and instruct to

engage student (IGS) (Hypothesis 2). It was noteworthy that the model detected a significant negative association between the main predictor of the study (CIT) and the two mediators, ACS $\beta = -0.22$ ($p < .05$) and IGS $\beta = -0.61$ ($p < .001$.) Statistically, this association means that teachers emphasis on the academic success (ACS) and their instruction to engage their student will be decreased by 0.22 SD and 0.61SD respectively when their collaboration increase by 1 SD, holding all else in the model constant. Another prominent finding was the negative and significant association between IGS and SES $\beta = -0.32$ ($p < .001$). This finding implies that teachers experienced difficulties about of 0.23 SD to engaging their student in the learning associated with 1SD increase in the SES holding all else in the model constant. As expected, student confidence in learning mathematics (CON) was associated positively and significantly to their engagement in learning mathematics (ENG) $\beta = 0.78$ ($p < .001$). Finally and unexpectedly, the model did not detect any effects for the leadership activities on different factors in both levels.

CHAPTER 5

DISCUSSION

The purpose of the present study was to obtain a better understanding of the effects of teacher collaboration in students' outcomes. Its conceptual framework suggested that close collaboration between teachers who share common purposes would help give them flexibility in adapting reforms to their local circumstances, which in turn would improve students' learning as the goal of any school reform. In line with the research questions, the present study utilized SEM to empirically delve more deeply into the direct relationship between teacher collaboration (CIT) and their fourth-grade students' engagement (ENG), confidence in learning (CON), and their mathematics achievement scores (PVM). Further, it investigated these relationships indirectly through two other teaching factors, emphasis on academic success (ACS) and instructing to engage student in learning (IGS). In this chapter, findings related to the study hypotheses will be discussed and connected to existing literature, particularly with respect to the significant findings that were identified in the previous chapter. In addition, implications of the study, its limitations, and recommendations for future research will be suggested.

5. 1 First Hypothesis: The Direct Relationship of CIT and ENG, CON, and PVM

The study hypothesized that teacher collaboration (school/ teacher level) positively and significantly predicts student engagement (ENG), confidence in learning (CON), and their mathematics achievement scores (student level). Surprisingly, no significant relationships were found. Hence, the hypotheses that the higher teacher collaboration as operationalized in this study will be associated with higher level of student achievement score in mathematics, their engagement, and confidence in learning was not supported by this study's data. This finding

corroborated with findings from previous studies particularly those of TIMSS data. Sturman and Lin (2011), for instance, except for a small effect in England's student performance, did not detect any other influence of the two items of TIMSS's collaboration measure (CIT), the main predictor of the current study, on eighth grade mathematics performance across Chinese Taipei, Republic of Korea, Singapore, Hong Kong, and Japan. It also echoes a recent study finding that did not reveal any association of the five indicators of CIT and student achievement in Japan (Reeves, Pun, & Chung, 2017).

One possible explanation of this finding is that Saudi teachers were engaged in a weak and contrived collaboration. While researchers asserted that deep level teacher collaboration seemed to be less frequent as teachers often tend to restrict it to practical affairs (Vangrieken et al., 2015), others warned against overly comfortable collaboration that was too weak and informal and that limited itself to contrived and superficial talk when exchanging materials and ideas (Hargreaves & Connor, 2017). Several factors could lead to this weak collaboration. At the individual level, this collaboration could be nourished by the tradition of isolation of teaching in Saudi Arabia as egg crates that focus on individual teacher responsibility and autonomy. In such culture, collaboration is less likely to be considered a systemic process in which teachers work together interdependently to impact their professional practice and student learning (Dufour, 2004; Wiseman, Alromi, & AlSadaawi, 2008). At the group level, weak collaboration would emerge as a consequence of groupthink and balkanization, which is a new form of isolation at the group level (Vangrieken et al., 2015). In such activities, teachers are more likely to engage in storytelling and helping activities that are less likely to lead to professional learning. A third catalyst of this of contrived collaboration between Saudi teachers could be at the structural level of educational system. After decades of top-down, bureaucratically-imposed educational reforms

(Alnahdi & Arabia, 2014; Chapman & Miric, 2009) with a limited participation from teachers in developing reform initiatives, teachers might feel that their voices in decision-making are limited. Therefore, they may only join superficially in what is seen as forced collaboration. Indeed, teachers in Saudi Arabia are required by their evaluation policy as a part of their professional development to make periodic reports on mutual classroom visitations (Ministry of Education, 1999). During these visits, teachers seem to engage in discussion, sharing lesson planning; however, their collaboration remained at the surface to avoid changing their underlying beliefs, which would inevitably lead to disagreement and conflict. This weak collaboration could be attributed to teachers' workload and their out-of-class time. Time pressures have been identified as a universal impediment to teacher collaboration in most educational systems (Hargreaves & O'Connor, 2017). Indeed, Al Tayyar (2014) found that some teachers were overworked by having to teach for 24 hours a week, the fulltime load work of teachers, which may have affected their ability to maintain a quality of teaching. He added that most teachers were dissatisfied with having to do organizational and administrative tasks, which they consider unrelated to their teaching and learning. These heavy workloads are more likely to leave teachers with no time for professional collaboration. Indeed, researchers underscored that the most viable time for teacher to learn is when they get outside their own classrooms and connect with each other (Hargreaves, 2001; OECD, 2015). Professional barriers could be another possible explanation of this weak collaboration. Jao and McDougall (2016) found that the lack of training or professional preparation for teachers to communicate and work with colleagues could deprive teachers of learning opportunities brought about through the context of collaboration. This situation could be exacerbated by the low participation of Saudi teachers in professional

development programs. Dodeen et al. (2012) found that only 29% of Saudi mathematics teachers participated in professional development programs in the previous two years.

Another possible interpretation of the lack of direct relationship between teacher collaboration and student outcomes could be attributed to the involved constructs' measurement. An effective teacher collaboration that can lead to professional learning might require more and deeper items to be captured than what occurs through the CIT. This interpretation can be supported because most studies that incorporated CIT measure did not find it to be a good predictor (Reeves et al., 2017; Sturman & Lin, 2011). Studies that utilized a more comprehensive measure of this concept were able to capture its effectiveness (Goddard et al., 2015; Miller et al., 2010; Ronfeldt et al., 2015). Measurement issues were not only for the predictor (CIT), but also for the outcome measures (ENG and CON). As shown in table 3 in the previous chapter, these measures suffered for method and translation effects (Abu-Hilal et al., 2014, 2013; Hooper et al., 2013; Nilsen, Trude, & Gustafsson, 2016). In addition, the of establishing a relationship between CIT and mathematics achievement scores (PVM) could be attributed to the lack of variance among fourth-grade students scores in this data. Actually, the average performance of Saudi fourth-grade was too low (410) when compared to the TIMSS 2011 centerpoint that was a score of 500. This excessively low performance could have exacerbated the measurement problems masking any school-level impac related to differences on the predictor (CIT).

5. 2 Second Hypothesis: The Mediation Effects of ACS and IGS on the Relationship between CIT and ENG, CON, and PVM

The study hypothesized that the effect of teacher collaboration (CIT) on student engagement (ENG), confidence in learning (CON), and their mathematics achievement scores is mediated by teacher emphasis on academic success (ACS) and their instruction to engage student

in learning (IGS). The study did not detect any indirect significant relationships between teacher collaboration and student outcomes. Hence, this hypothesis was not supported by this study's data. This finding is partly in line with the findings of Blömeke, Vegar, and Suhl (in Nilsen & Gustafsson, 2016) that IGS, among other several predictors, was not a good predictor of student achievement in mathematics for fourth grade across 47 countries. This finding is also echoed other recent finding by Nilsen and Gustafsson (2016) that an emphasis on academic success, ACS in the current study, did not predict student mathematics achievement scores. When they examined its items individually, only two items that reflect parental support for student achievement and students' desire to perform well were significantly related to student achievement.

As aforementioned, measurement issues could be one reason for the lack of effectiveness of these two mediators. As shown in Table 3 in the measurement model of this study, both measures did not fit the data; several modifications were conducted to reach a good fit. This last finding was also supported by Nilsen and Gustafsson (2016) when they concluded that SEAS did not satisfy ideals of unidimensionality. Another possible explanation of the ineffectiveness of ACS and IGS as mediators of CIT on PVM could be attributed to culture differences. Ker (2015) caution against such cultural differences, particularly in ILSAs studies such as TIMSS. While TIMSS measures have been developed within a well conceptual framework, they have been selected and defined mostly in Western cultures. In countries such as Saudi Arabia, some of these measures such as ACS and IGS were not straightforward concepts. Therefore, a key measure is likely to be effective in one educational system but not essential in another.

Another noteworthy finding in the mediation relationships is the negative association between CIT and both mediators, ACS and IGS. This negative association, however, should not

be interpreted as supporting a conclusion that professional teacher collaboration is of no importance, but instead it should be interpreted as indicating a need for further research. Two explanations can be offered for this finding. First, there is a possibility that the collective different barriers that buffered teacher collaboration (CIT) in the first place were quite likely pushing for a long time for the wrong kinds of collaboration in the wrong way. Researchers caution against the harmful view of teacher collaboration that would reduce teachers' motivation to initiate it in the first place (Datnow, 2011; Hargreaves & O'Connor, 2017; Vangrieken et al., 2015). In fact, Hargreaves (1994) stated that "Collaboration and restructuring can be helpful or harmful, and their meanings and realizations therefore need to be inspected repeatedly to ensure that their educational and social benefits are positive" (p. 248). This interpretation could lead to the second explanation of this negative finding that can be obtained from the statistical perspective. In correlational data such the current study's data, reversed causality could be a possible reason. This well-known problem in educational research, particularly in cross-sectional studies, concerns the causal mechanisms of variables (Nilsen & Gustafsson, 2016). Therefore, it could be reasonable that teachers did not appreciate a weak and contrived collaboration as much as they emphasized their students' academic success and engaging them in learning. Thus, the more they spent their time focusing on the ACS and IGS, the less time they allocate for such a weak collaboration.

Another notable finding is the negative relationship between SES and IGS in the school level. This negative association means that the higher SES of the students, the more difficult it is for their teachers to engage them in learning. Two explanations could help interpreting this finding. First, it could be attributed to measures of these two constructs. As shown in Table 3 in the measurement model, both measures suffered from misfit, particularly the SES measure. This

measure had 11 items, and out of these, 6 items were left to participating countries and provinces to be added based on how they conceptualized their study's population SES. In the case of Saudi Arabia's TIMSS 2011 data, most items that were added were, unfortunately, a source of misfit. Although one of these items was deleted, three other items were poorly loaded. The other explanation can be obtained from the school effectiveness perspective. Since this relationship was found only to be significant in school level, it indicates that there were significant differences between schools related to their students' SES, which is consistent with literature since Coleman's report in 1966 (Baker et al., 2002; Coleman, 1966). Second, this negative association should not be interpreted as supporting a conclusion that high student SES could lower schools' effectiveness. Instead, it should be interpreted as indicating a need for further research. It could also be interpreted that teaching quality, here the IGS, should be supported to meet different students' instructional needs. Indeed, Mullis, Martin, and Ina (2011) asserted that teacher preparation is a powerful factor in student learning that could overcome socioeconomic background factors. This finding echoed Akiba et al.'s (2007) work that found a gap between high and low SES students in their access to qualified teachers. However, this gap was not found to be significant in explaining the difference in student achievement scores in mathematics. They operationalized teacher quality based on credentials and teaching experience. Thus, it is reasonable to assume it could overcome the differences in students' SES. This latter interpretation could be supported by the conclusion of Akiba et al. (2007) that more attention should be paid to important mediators that lead to effective instructional practice among all teachers.

A final substantial finding of this study is the negative relationship between student confidence in learning (CON) and their mathematics achievement scores (PVM) in student level.

This finding echoed previous studies that used TIMSS data (Loveless & Diperna, 2006; Zhao, 2012, 2017). This finding suggests that focusing on rising only test scores would scarify other psychological aspects that are necessary for student learning such as their confidence in learning. Surprisingly, at the school level, this negative relationship did not appear ($\beta = -0.612$ ($p < 0.14$), which is inconsistent with the previous studies (Marsh et al., 2015). One explanation of this finding could be attributed to the measure of this concept. While the current study was able to establish a good fit of this measure with six items as shown in the measurement model, other researchers suggested that this measure consists of two strongly correlated dimensions, self-concept and subject value; however, they are related to different subdomains of academic achievement (Abu-Hilal et al., 2014). Another interpretation could be that evaluative items, for example, “I usually do well in math” and “math is more difficult for me than peers,” are not good measures of self-concept among Arab students. Abu-Hilal et al. (2014) stated that “school system and the family structure do not encourage children to be critical, make decisions, and shoulder responsibility while they are children and adolescents” (279). Therefore, it would be confusing to ask students to evaluate their abilities. This situation could be exacerbated with young students such the current study’s sample. This finding indicates, however, that more research is needed in this area.

5. 3 Implications

In the Saudi educational system, top-down reforms have been the prevalent way (Alnahdi & Arabia, 2014; Chapman & Miric, 2009). To challenge this type of educational reform, the current study was conceptualized in a framework suggested by Tyack and Cuban’s (1995) that bringing “about improvement at the heart of education, classroom instruction, will not be achieved unless an internal change created by the knowledge and expertise of teachers are

facilitated and supported” (p. 135). Therefore, the study examined two premises. First, the study hypothesized that students’ outcomes (ENG, CON, PVM) are significantly and positively associated with their teachers’ collaboration. Second, it hypothesized that this association is mediated by teachers’ emphasis on academic success (ACS) and instruct to engage student in learning (IGS). However, none of these hypotheses were supported by the data. The results suggested that teacher collaboration, if it is not properly structured, might not contribute to improved student outcomes, as measured by student engagement, confidence in learning, and their achievement scores in mathematics. Most prominently, the study found that teacher collaboration could not only be ineffective, but it could also be harmful. The study detected a negative relationship between teacher collaboration, as operationalized and measured in this study, emphasis on the academic success (ACS), and engaging student in learning (IGS) as reported by teachers.

These findings have several implications for Saudi educational policy makers and practitioners alike. First, although this study did not show any connection between teacher collaboration and student outcomes, several previous studies have established that connection (Goddard, Yvonne L. Goddard, Roger D. Tschannen-Moran, 2007; Goddard et al., 2015; Miller et al., 2010). Therefore, rather than dictating from the top, policy makers should collaboratively work with practitioners in schools to create conditions that may lead to an effective collaboration between teachers which can improve school environments and eventually students outcomes. However, policy makers should also be aware that establishing the structures and cultures of interdependency collaboration at schools takes time and efforts and should not be rushed towards quick win initiatives. Third, although the current study utilized a cross-sectional data, the literature suggests a continuum of teacher collaboration ranging from weak form of collegiality

to joint work (Hargreaves & O'Connor, 2017). Thus, it is reasonable to assume that teacher collaboration this study examined was in the beginning of the continuum, a weak one, and because of the flawed initiation combined with inappropriate implementation, it flipped furiously to the negative side. Therefore, it is not an adequate for policy makers to just initiate teacher collaboration initiatives, but also work jointly with practitioners during the implementation phases and periodically revisit these initiatives to ensure that they are still in the positive direction. Fourth, the study found that student confidence in learning mathematics was a good predictor of student engagement in learning mathematics in both within and between schools. Therefore, educators and policy makers in particular should consider these student outcomes beside test scores. Fifth, the study found that student socioeconomic statuses negatively predicted teacher instruction to engage student (IGS). While this finding requires more research towards deep understanding of this negative relationship, it also requires that educators and school leaders should work collaboratively with teachers and support them to meet their student needs. Sixth, the main purpose of ILSAs such TIMSS is to periodically provide reliable data that can offer educational policy makers and educators a better understanding of their educational system in general and in particular point to promising instructional practices that could support reform classroom teaching (Mullis, Martin, Goh & Cotter, 2016). Capitalizing in these ILSAs, however, depends on the quality of data collection. As this current study found, several measures of TIMSS 2011 that were used to collect Saudi data were poorly translated. Moreover, other significant measures such as pre-schooling seemed to be neglected. Furthermore, the SES measure imprecisely was included several items that appeared unrelated. Therefore, Saudi policy makers should exert more efforts to make sure that participating in these ILSAs is paying off by examining the validity of their data collection instruments. Seventh, although TIMSS is

fundamentally an achievement study, it does not include a prior attainment. Thus, it might be prudent for Saudi policy makers to consider aligning another achievement measure with TIMSS data that can be used to control for the prior attainment that can lead to more accurate findings and interpretations. Finally, applying statistical procedures such as MSEM requires quite large samples (Byrne, 2012). Thus, it could be more practical for Saudi policy makers to increase their student, classroom, and school samples in the ILSAs such as TIMSS and PIRLS.

5. 4 Limitations and Future Research

Whilst this study provided several empirical findings, there are also limitations that need to be noticed. First, this study was based on TIMSS data that is by nature cross sectional and can only provide correlations between variables. Thus, causality cannot be inferred based on such data. To strength causal inference, future research could capitalize on a longitudinal data. Second, concerns could be raised about the reliability and validity of teacher assessments of their own instructional practices. Social desirability bias in self-report assessments is a well-known threat to validity in social science research. Thus, qualitative studies could offer more understanding of the nature of teacher collaboration and its relationship to other teacher and student variable. Third, the study findings detected several validity threats to TIMSS measures. Therefore, investigating psychometric properties of TIMSS measures that involve Saudi samples seem essential. Fourth, given the unpredictable finding of the negative relationship between teacher collaboration and their instructional practices (ACS& IGS), future studies that use TIMSS 2011 of eighth grade instead of fourth grade and probably other TIMSS data cycles, could offer more understanding of such results. Fifth, the study did not capture any effects of teacher collaboration in student mathematics outcomes. Examining these relations utilizing science (TIMSS) and reading (PIRLS) data is more likely to provide a better understanding of

the relation between these variables. Sixth, the hypothesized mediators of the study were found to be unrelated to student outcomes. Future studies of TIMSS should pay attention to other mediators that could lead to effective student outcomes. Finally, considering the vital role of teacher collaboration as an internal change agent of educational reform, constructing and validating a more comprehensive measure that can capture more quantity and quality activities of teacher collaboration is fundamental.

REFERENCES

- Abu-Hilal, M., Al Hussain, A. (1997). Dimensionality and Hierarchy of the Sdq in a Non-Western Milieu A Test of Self-Concept Invariance Across Gender. *Journal of Cross-Cultural Psychology*, 28(5), 353-553.
- Abu-Hilal, M., Abdelfattah, F. A., Alshumrani, S. A., Abduljabbar, A. S., & Marsh, H. W. (2013). Construct validity of self-concept in TIMSS's student background questionnaire: A test of separation and conflation of cognitive and affective dimensions of self-concept among Saudi eighth graders. *European Journal of Psychology of Education*, 28(4), 1201–1220. doi: 10.1007/s10212-012-0162-1
- Abu-Hilal, M. M., Abdelfattah, F. A., Shumrani, S. A., Dodeen, H., Abduljabber, A. S., & Marsh, H. W. (2014). Mathematics and science achievements predicted by self-concept and subject value among 8th grade Saudi students: Invariance across gender. *International Perspectives in Psychology: Research, Practice, Consultation*, 3(4), 268–283. doi: 10.1037/ipp0000022
- Aburizaizah, S., Kim, Y., & Fuller, B. (2016). Diverse schools and uneven principal leadership in Saudi Arabia. *International Journal of Educational Research*, 80, 37-48. doi: 10.1016/j.ijer.2016.08.007
- Akiba, M. (2012). Professional learning activities in context: A statewide survey of middle school mathematics teachers. *Education Policy Analysis Archives*, 20(14), 1-31.
- Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, 36(7), 369-387. doi: 10.3102/0013189X07308739
- Al Issa, A. (2009). *Education Reform in KSA: Lack of Political Vision, Religious Cultural Obsession, and Educational Management Failure*. Dar Al Saqi Publishing, Beirut.
- Alnahdi, G. H., & Arabia, S. (2014). Educational change in Saudi Arabia. *Journal of International Education Research*, 10(1), 1-7.
- Al Otaibi, Badr. (2008). Facing the reform of secondary education in the Kingdom of Saudi Arabia in the light of contemporary global trends. *Journal of the College of Education*, King Saud University, April (14), 415-482.
- Alreshidi, N. A. K. (2016). *Investigating problem-based learning in Saudi Arabian mathematics education: a TIMSS-related study* (Doctoral dissertation, University of Glasgow).
- Alsadaawi, A. S. (2010). Saudi national assessment of educational progress (SNAEP). *Journal of Education Policy*, 5(11), 1–14.
- Al Tayyar, K. (2014). *Job satisfaction and motivation amongst secondary school teachers in Saudi Arabia* (Doctoral dissertation, University of York).
- Alyami, R. H. (2014). Educational Reform in the Kingdom of Saudi Arabia : Tatweer Schools as a Unit of Development, *Literacy Information and Computer Education Journal (LICEJ)*, 5(2), 1424 –1433.
- Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the “Heyneman-Loxley effect” on mathematics and science achievement. *Comparative education review*, 46(3), 291-312.
- Banerjee, N., Stearns, E., Moller, S., & Mickelson, R. A. (2017). Teacher job satisfaction and student achievement: The roles of teacher professional community and teacher collaboration in schools. *American Journal of Education*, 123(2), 1-39.

- Bartalo, D. B. (2012). *Closing the teaching gap: Coaching for instructional leaders*. Corwin Press.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258.
- Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional Community in Chicago Elementary Schools: Facilitating Factors and Organizational Consequences. *Educational Administration Quarterly*, 35(5), 751–781. <http://doi.org/10.1177/0013161X99355004>
- Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. University of Chicago Press.
- Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.
- Cai, E. Y. L., & Liem, G. A. D. (2017). ‘Why do I study and what do I want to achieve by studying?’ Understanding the reasons and the aims of student engagement. *School Psychology International*, 38(2), 131-148.
- Caponera, E., & Losito, B. (2016). Context factors and student achievement in the IEA studies: Evidence from TIMSS. *Large-Scale Assessments in Education*, 4(1), 1-22. <http://doi.org/10.1186/s40536-016-0030-6>
- Chapman, D. W., & Miric, S. L. (2009). Education quality in the Middle East. *International Review of Education*, 55, 311–344. doi: 10.1007/s11159-009-9132-5
- Coke, P. K. (2005). Practicing what we preach: An argument for cooperative learning opportunities for elementary and secondary educators. *Education*, 126, 392–398. <http://search.proquest.com/docview/196441472?accountid=17215>
- Coleman, J. S. (1966). *Equality of Education Opportunity Study*, Washington, DC: US Department of Health, Education, and Welfare, Office of Education. National Center for Education Statistics.
- Cooper, C., and J. Boyd. 1997. A caring, competent and continuously improved teacher for every child in Australia: Dream or reality? *Global Learning Community*, Available online: http://www.vision.net.au/~globallearning/pages/lfs/cb_article.html
- Crockett, S. A. (2012). A five-step guide to conducting SEM analysis in counseling research. *Counseling Outcome Research and Evaluation*, 3(1), 30–47. doi: 10.1177/2150137811434142
- Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890-1990*. Teachers College Press.
- Cuban, L. (1996). Myths about changing schools and the case of special education. *Remedial and Special Education*, 17(2), 75-82.
- Cuban, L. (1995). The hidden variable: How organizations influence teacher responses to secondary science curriculum reform. *Theory into Practice*, 34(1), 4-11.
- Curry, M. (2008). Critical friends groups: The possibilities and limitations embedded in teacher professional communities aimed at instructional improvement and school reform. *Teachers College Record*, 110(4), 733-774.
- Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should know and be able to do*. San Francisco, CA: JosseyBass
- Darling-Hammond, L., & Sykes, G. (Eds.). (1999). *Teaching as the learning profession: Handbook of policy and practice*. San Francisco: Jossey-Bass.
- Darling-Hammond, L., & Wei, R. C. (2009). *Professional learning in the learning profession : A*

- status report on teacher development in the United States and abroad*. National Staff Development Council, 1–32. doi: 10.1006/jfbi.2002.2063
- Datnow, A. (2011). Collaboration and contrived collegiality: Revisiting Hargreaves in the age of accountability. *Journal of educational change*, 12(2), 147-158.
- Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental health agencies. *Journal of Emotional and Behavioral Disorders*, 19(1), 27–40. doi: 10.1177/1063426610365879
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1), 195–210. doi: 10.1162/003465304323023750
- Devaney and Sykes (1988). Making the Case of Professionalism. In Lieberman, A. (Ed.), *Building a professional culture in schools*. Teachers College Press, Teachers College, Columbia University, 1234 Amsterdam Ave., New York, NY 10027.
- Dodeen, H., Abdelfattah, F., Shumrani, S., & Hilal, M. A. (2012). The effects of teachers' qualifications, practices, and perceptions on student achievement in TIMSS mathematics: A comparison of two countries. *International Journal of Testing*, 12(1), 61–77. doi: 10.1080/15305058.2011.621568
- Dufour, R. (2004). What is a professional learning community? *Educational Leadership*, 61(8), 6–11. Retrieved from http://www.ascd.org/publications/educational_leadership/may04/vol61/num08/What-Is-a-Professional-Learning-Community.aspx
- Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, 16, 149–167. doi: 10.1016/j.leaqua.2004.09.009
- Forte, A. M., & Flores, M. A. (2014). Teacher collaboration and professional development in the workplace: A study of Portuguese teachers. *European Journal of Teacher Education*, 37(1), 91–105. doi: 10.1080/02619768.2013.763791
- Foy, P. (2013). *TIMSS and PIRLS 2011 User Guide for the Fourth Grade Combined International Database*. Boston College.
- Foy, P., Galia, J., & Li, I. (2008). *Scaling the TIMSS 2007 mathematics and science assessment data*. In J.F. Olson, Martin, M.O & I.V.S. Mullis (Eds.), TIMSS 2007 technical report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Foy, P., & Olson, J. F. (Eds.). (2009). *TIMSS 2007 international database and user guide*. TIMSS & PIRLS International Study Center, Boston College.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- Goddard, R., Goddard, Y., Kim, E. S., & Miller, R. (2015). A theoretical and empirical analysis of the roles of instructional leadership, teacher collaboration, and collective efficacy beliefs in support of student learning. *American Journal of Education*, 121(4), 501–530.
- Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers college record*, 109(4), 877-896.
- Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2001). *Supervision and instructional leadership: A developmental approach*. Allyn & Bacon/Longman Publishing, a Pearson Education Company, 1760 Gould Street, Needham Heights, MA 02494. Web site: <http://www. abacon. com>
- Gustafsson, J.-E. E., Nilsen, T., & Hansen, K. Y. (2015). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8.

- Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*. doi: 10.1016/j.stueduc.2016.09.004
- Hackmann, D. G., Petzko, V. N., Valentine, J. W., Clark, D. C., Nori, J. R., & Lucas, S. E. (2002). Beyond interdisciplinary teaming: Findings and implications of the NASSP national middle level study. *NASSP Bulletin*, 86(632), 33-47.
- Haberman, M. (2004) Can star teachers create learning communities? *Educational Leadership*, 61(8), 52-56
- Hargreaves, A. (1994). *Changing teachers, changing times: Teachers' work and culture in the postmodern age*. New York, NY: Teachers College Press
- Hargreaves, A., & Fullan, M. (2013). The Power of Professional Capital: With an Investment in Collaboration, Teachers Become Nation Builders. *Journal of Staff Development*, 34(3), 36-39.
- Hargreaves, A., & O'Connor, M. (2017). Cultures of professional collaboration: Their origins and opponents. *Journal of Professional Capital and Community*, 2(2), 74–85. doi: 10.1108/JPC-02-2017-0004
- Hargreaves, A., & Connor, M. (2017). *Collaborative Professionalism*. The World Innovation Summit for Education. Qatar.
- Hargreaves, D. H. (2001). A capital theory of school effectiveness and improvement [1]. *British educational research journal*, 27(4), 487-503.
- Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development*, 65(6), 1541–1545.
- Hooper, M., Arora, A., Martin, M. O., & Mullis, I. V. S. (2013). Examining the behavior of “reverse directional” items in the TIMSS 2011 Context Questionnaire Scales. In *IEA International Research Conference*. Nanyang, Singapore.
- Honingh, M., & Hooge, E. (2014). The effect of school-leader support and participation in decision making on teacher collaboration in Dutch primary and secondary schools. *Educational Management Administration & Leadership*, 42(1), 75-98.
- Holmes, D. H., Futrell, M. H., Christie, J. L., & Cushman, E. J. (1995). School Restructuring: Meeting Teachers' Professional Development Needs. In *annual meeting of the American Educational Research Association, San Francisco*.
- Hord, S. M. (1997). Professional learning communities: Communities of continuous inquiry and improvement. *Leadership*, 40(1), 58–59. doi: 10.1177/1365480210376487
- House, J. D., & Telese, J. A. (2015). Engagement in science lessons and achievement test scores of eighth-grade students in Korea: Findings from the TIMSS 2011 assessment. *Education*, 135(4), 435–438.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi: 10.1080/10705519909540118
- Huang, F. L., & Cornell, D. G. (2016). Using multilevel factor analysis with clustered data: Investigating the factor structure of the positive values scale. *Journal of Psychoeducational Assessment*, 34(1), 3–14. doi: 10.1177/0734282915570278
- Jao, L., & McDougall, D. (2016). Moving beyond the barriers: Supporting meaningful teacher collaboration to improve secondary school mathematics. *Teacher Development*, 45(30), 1–17. doi: 10.1080/13664530.2016.1164747
- Johnson, S. M., Berg, J. H., & Donaldson, M. L. (2005). *Who stays in teaching and why?: A review of the literature on teacher retention*. The Project on the Next Generation of

Teachers. Harvard Graduate School of Education, 1–134.

- Kelchtermans, G. (2006). Teacher collaboration and collegiality as workplace conditions. A review. *Zeitschrift Für Pädagogik*, 52(2), 220–237. Retrieved from http://www.pedocs.de/frontdoor.php?source_opus=4454
- Ker, H. W. (2015). The impacts of student-, teacher- and school-level factors on mathematics achievement: An exploratory comparative investigation of Singaporean students and the USA students. *Educational Psychology*, 36(2), 254–276. doi: 10.1080/01443410.2015.1026801
- Kloosterman, P. (1988). Self-confidence and motivation in mathematics. *Journal of Educational Psychology* 80(3), 345–351.
- Koyame-Marsh, R. O. (2017). The Dichotomy Between the Saudi Women's Education and Economic Participation. *The Journal of Developing Areas*, 51(1), 431-441.
- Kung, H. Y. (2009). Perception or confidence? self-concept, self-efficacy and achievement in mathematics: A longitudinal study. *Policy Futures in Education*, 7(4), 387–398. doi: 10.2304/pfie.2009.7.4.387
- Leithwood, Kenneth; Seashore, Karen; Anderson, Stephen; Wahlstrom, Kyla; Center for Applied Research and Educational Improvement. (2004). *Review of Research: How Leadership Influences Student Learning*. University of Minnesota, Center for Applied Research and Educational Improvement. Retrieved from the University of Minnesota Digital Conservancy, <http://hdl.handle.net/11299/2035>.
- Leonard, L., & Leonard, P. (2003). The continuing trouble with collaboration: Teachers talk. *Current Issues in Education*, 6.
- Loveless, T., & Diperna, P. (2000). *How well are American students learning?*. Brown Center on Education Policy, The Brookings Institution.
- Liang, G., Zhang, Y., Huang, H., Shi, S., & Qiao, Z. (2015). Professional development and student achievement: International evidence from the TIMSS data. *Journal of Postdoctoral Research*, 3(2), 17-31.
- Lieberman, A. (1988). *Building a professional culture in schools*. Teachers College Press, Teachers College, Columbia University, 1234 Amsterdam Ave., New York, NY 10027.
- Lima, J. A. (2001). Forgetting about friendship: Using conflict in teacher communities as a catalyst for school change. *Journal of Educational Change*, 2,97–122. doi:10.1023/A:1017509325276.
- Little, J. (2013). Multilevel confirmatory ordinal factor analysis of the life skills profile-16. *Psychological Assessment*, 25(3), 810–25. doi: 10.1037/a0032574
- Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American educational research journal*, 19(3), 325-340.
- Little, J. W. (1999). Organizing Schools for Teacher Learning. In Darling-Hammond, L., Sykes, G. (Ed.), *Handbook of Policy and Practice: Teaching As the Learning Profession* (pp. 233-262). *Jossey-Bass Education Series*. Jossey-Bass Inc., Publishers, 350 Sansome St., San Francisco, CA 94104.
- Little, J.W. (1999). Teachers professional development in the context of high school reform: Findings from a three-year study of restructuring schools. *Paper presented at the Annual Meeting of the American Educational Research Association*, Montreal, April.

- Little, J. W. (1992). "The black box of professional community." In A. Lieberman (ed.), *The Changing Contexts of Teaching* [Ninety- First yearbook of the National Society for the Study of Education, Part 1] (pp. 157-178). Chicago: University of Chicago Press.
- Little, J. W., Brodeur, M., Deaudelin, C., Bru, M., Gameda, F. T., Fiorucci, M., ... Lampert, M. (2009). Learning teaching in, from, and for practice: What do we mean? *Teaching and Teacher Education*, 25(1), 21–34. doi: 10.1177/0022487109347321
- Little, J. W. (1987). Teachers as colleagues. In V. Richardson-Koehler (Ed.), *Educators' handbook: A research perspective* (pp. 491-518). New York: Longman
- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, 15(2), 129–151.
- Lohman, M. C. (2005). A survey of factors influencing the engagement of two professional groups in informal workplace learning activities. *Human Resource Development Quarterly*, 16(4), 501-527.
- Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). Professional communities and student achievement – a meta-analysis. *Teaching and Teacher Education*, 27(4), 722–731. doi: 10.1016/j.tate.2010.12.003
- Louis, K.S. & Kruse, S.D. (1995). *Professionalism and community: Perspectives on reforming urban schools*. Thousand Oaks, CA: Corwin Press.
- Maroun, N., Samman, H., Moujaes, C., & Abouchakra, R. (2008). *How to succeed at education reform: The case for Saudi Arabia and the broader GCC region*. Abu Dhabi, Ideation Center, Booz & Company, (June), 109–113. Retrieved from http://www.theideationcenter.com/media/file/How_to_Succeed_at_Education_Reform_FIN_AL.pdf
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., ... Parker, P. (2013). Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, 105(1), 108–128. doi: 10.1037/a0029907
- Marsh, H. W., Abduljabbar, A. S., Morin, A. J. S., Parker, P., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. M. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology*, 107(1), 258–271. doi: 10.1037/a0037485
- Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., & Nagengast, B. (2014). The Big-Fish-Little-Pond Effect in Mathematics: A Cross-Cultural Comparison of U.S. and Saudi Arabian TIMSS Responses. *Journal of Cross-Cultural Psychology*, 45(5), 777–804. <http://doi.org/10.1177/0022022113519858>
- McLaughlin, M. W., & Marsh, D. D. (1990). Staff development and school change. *Teachers College Record* (Vol. 80). Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=1091>
- Miles, M. B., Saxl, E. R., & Lieberman, A. (1988). What skills do educational “change agents” need? An empirical view. *Curriculum Inquiry*, 18(2), 157-193.
- Miller, R. J., Goddard, Y. L., Goddard, R., Larsen, R., & Jacob, R. (2010). Instructional leadership: A pathway to teacher collaboration and student achievement, *Educational Administration Convention*, New Orleans, LA.

- Ministry of Education. (1999). *Educational Supervisors' Guidance*. General Directory of Supervision. First edition.
- Moolenaar, N. M., Daly, A. J., & Slegers, P. J. (2010). Occupying the principal position: Examining relationships between transformational leadership, social network position, and schools' innovative climate. *Educational administration quarterly*, 46(5), 623-670.
- Moolenaar, N. M., Slegers, P. J., & Daly, A. J. (2012). Teaming up: Linking collaboration networks, collective efficacy, and student achievement. *Teaching and Teacher Education*, 28(2), 251-262.
- Morimoto, K. (1973). Notes on the context for learning. *Harvard Educational Review*, 43(2), 245-257.
- Morse, W. C. (2000). Foreword. In M. Friend & L. Cook, *Interactions: collaboration skills for school professionals* (3rd ed., pp. xi-xii). New York: Addison Wesley Longman
- Moyer, P. S., & Milewicz, E. (2002). Learning to question: Categories of questioning used by preservice teachers during diagnostic mathematics interviews. *Journal of Mathematics Teacher Education*, 5(4), 293-315.
- Mullis, I. V., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004). *International student achievement in mathematics*. TIMSS 2003 International Mathematics Report, 31-54.
- Mullis, I. V., & Martin, M. O. (2011). TIMSS and PIRLS 2011: *Relationships Among Reading, Mathematics, and Science Achievement—Implications for Early Learning*. TIMSS and PIRLS, 1-11.
- Mullis, I. V. S., Martin, M. O., Goh, S., & Cotter, K. (Eds.) (2016). *TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2015/encyclopedia/>
- Musca, S. C., Kamiejski, R., Nugier, A., Méot, A., Er-Rafiy, A., & Brauer, M. (2011). Data with hierarchical structure: Impact of intraclass correlation and sample size on Type-I error. *Frontiers in Psychology*, 74(2), 1–6. doi: 10.3389/fpsyg.2011.00074
- Muthén, B. O., Kao, C. F., Burstein, L., Muthen, B. O., Kao, C. F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement*, 28(1), 1–22. doi: 10.1111/j.1745-3984.1991.tb00340.x
- Muthén, L. K., & Muthén, B. O. (1998-2015). *MPlus (Version, 7.1). [Computer Software]*. Los Angeles, CA: Muthén & Muthén.
- Newmann, F. M. (1992). *Student Engagement and Achievement in American Secondary Schools*. New York, NY: Colleague Teachers Press.
- Nilsen, Trude; Gustafsson, J.-E. (2016). Teacher quality , instructional quality and student outcomes. AG, Switzerland: International Association for the Evaluation of Educational Achievement (IEA).
- OECD. (2015). *Embedding professional development in schools for teacher success*. Teaching in Focus, 10. Paris: OECD Publishing. <http://dx.doi.org/10.1787/5js4rv7s7snten>.
- Pendlebury, S. (1990). Practical arguments and situational appreciation in teaching. *Educational Theory*, 40(2), 171-179.
- Peterson, K. D.; Brietzke, R. (1994). *Building collaborative cultures seeking ways to reshape urban schools*. Oak Brooks, IL: North Central Regional Educational Laboratory.
- Peterson, P. L., McCarthey, S. J., & Elmore, R. F. (1996). Learning from School Restructuring. *American Educational Research Journal*, 33(1), 119–153. doi: 10.2307/1163383

- Reeves, P. M., Pun, W. H., & Chung, K. S. (2017). Influence of teacher collaboration on job satisfaction and student achievement. *Teaching and Teacher Education*, 67, 227–236. doi: 10.1016/j.tate.2017.06.016
- Richardson, V. (1998). *How teachers change: What will lead to change that most benefits student learning?* National Center for the Study of Adult Learning Literacy, 2(4), 7–11.
- Richardson, V. (1994). Conducting research on practice. *Educational Researcher*, 23(5), 5–10.
- Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American educational research journal*, 28(3), 559-586.
- Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, 52(3), 475-514. doi: 10.3102/0002831215585562
- Rosenholtz, S. J. (1989). Workplace conditions that affect teacher quality and commitment: Implications for teacher induction programs. *The Elementary School Journal*, 89(4), 421-439. doi: 10.1086/461584
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142–151. doi: 10.3102/0013189X10363170
- Saudi Arabia's Vision 2030, 2016. *National Transformation Program 2020*. [Online]. Viewed 1 February 2018, <<http://vision2030.gov.sa/en/ntp>>
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools. A review of school effectiveness research*. London: OFSTED.
- Schleicher, A. (2012). Building a high-quality teaching profession. Lessons from around the world. *Educational Studies*, (1), 74-92.
- Schleicher, A. (2011). Lessons from the world on effective teaching and learning environments. *Journal of Teacher Education*, 62(2), 202-221.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–337.
- Schwartz, D., & Bransford, J. D. (2005). Efficiency and Innovation in Transfer. *Transfer of Learning from a Modern Multidisciplinary Perspective*, (3), 1–51. doi: 10.1111/j.1365-2133.2005.06492.x
- Scribner, J. P. (1999). Professional development: Untangling the influence of work context on teacher learning. *Educational Administration Quarterly*, 35(2), 238–266. doi: 10.1177/0013161X99352004
- Senge, P. (1990). *The fifth discipline: The art and science of the learning organization*. New York: Currency Doubleday.
- Smaller, H., Hart, D., Clark, R., Livingstone, D. (1997). *Informal/Formal Learning and Workload among Ontario Secondary School Teachers*. NALL Working Paper. Toronto, Ontario, Canada: Ontario Institute for Studies in Education.
- Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35(2), 87-100.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of educational change*, 7(4), 221-258. doi: 10.1007/s10833-006-0001-8

- Sturman, L., & Lin, Y. (2011). Exploring the mathematics gap: TIMSS 2007. *Ricerca Azione*, 3(1), 43–58.
- Swanson, C. B., & Barlage, J. (2006). Influence: A Study of the Factors Shaping Education Policy. *Editorial Projects in Education*.
- Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research*, 35(4), 1048–1079. <http://doi.org/10.1016/j.ssresearch.2005.08.001>
- King Abdullah Bin Abdul-Aziz Public Education Development Project (TATWEER) (2008). [Online]. Viewed 1 February 2018, <https://www.tatweer.edu.sa/>
- The Public Education Evaluation Commission (PEEC) (2016.) The objectives of PEEC. [Online]. Viewed 1 February 2018, <<http://www.peec.gov.sa/objectives-of-the-commission?lang=en>>.
- The World Bank, (2011). Fertility rate. Viewed 12 March 2017 from <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN>
- Tyack, D. B., & Cuban, L. (1995). *Tinkering toward utopia*. Harvard University Press.
- Tyack, D., & Tobin, W. (1994). The “grammar” of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31(3), 453-479
- UNESCO-IBE. (2011). *World Data on Education-Saudi Arabia*. Wde, 31. Retrieved from <http://www.ibe.unesco.org/>
- UNICEF. (2007). *An overview of child well-being in rich countries*. Report Card 7. Florence: UNICEF Innocenti Research Centre.
- Van Gasse, R., Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2016). Teacher collaboration on the use of pupil learning outcome data: A rich environment for professional learning?. *Teaching and Teacher Education*, 60, 387-397. doi: 10.1016/j.tate.2016.07.004
- Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review*, 15, 17–40. doi: 10.1016/j.edurev.2015.04.002
- Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80–91. doi: 10.1016/j.tate.2007.01.004
- Wang, J. (2001). TIMSS primary and middle school data: Some technical concerns. *Educational Researcher*, 30(6), 17–21. doi: 10.3102/0013189X030006017
- Waldron, N. L., & McLeskey, J. (2010). Establishing a collaborative school culture through comprehensive school reform. *Journal of Educational and Psychological Consultation*, 20(1), 58-74.
- Williams, M. L. (2010). *Teacher collaboration as professional development in a large, suburban high school* (Unpublished doctoral dissertation).
- Wiseman, A. W., & Alromi, Naif H, AlSadaawi, A. (2008). Educational indicators and national development in Saudi Arabia. *Paper presnted to IEA International Research Conference, September 18-20 2008*. Taipei City, Taiwan.
- Zhao, Y. (2017). What works may hurt: Side effects in education. *Journal of Educational Change*, 18(1), 1–19. doi: 10.1007/s10833-016-9294-4
- Zhao, Y. (2012). *World class learners: Educating creative and entrepreneurial students*. Corwin Press.
- Zito, M. F. (2011). *Is working together worth it?: Examining the relationship between the quality of teacher collaboration, instruction, and student achievement* (Doctoral dissertation, University of Massachusetts Amherst).

Appendixes

Appendix 1.

TIMSS student engagement in mathematics items:

- 1) I know what my teacher expects me to do.
- 2) I think of things not related to the lesson.
- 3) My teacher is easy to understand.
- 4) I am interested in what my teacher says.
- 5) My teacher gives me interesting things to do.

Appendix 2.

TIMSS student confidence in learning mathematics:

- 1) I usually do well in mathematics
- 2) Mathematics is harder for me than for many of my classmates*
- 3) I am just not good at mathematics*
- 4) I learn things quickly in mathematics
- 5) I am good at working out difficult mathematics problems
- 6) My teacher tells me I am good at mathematics
- 7) Mathematics is harder for me than any other subject.

Appendix 3.

TIMSS scales coding:

- Teacher collaboration ATBG10 (A, B, C, D, E),
- Teacher emphasis on academic success ATBG06 (B, C, D, E, H),

- Engage student in learning ATBG15 (A, B, C, D, E, F),
- Leadership activities ACBG15 (A, B, C, D, E, F, G, H, I, J, K, L),
- Student engagement in learning ASBS05
- Student confidence in learning ASBS06
- Student SES ASBG05 (A, B, C, D, E, F, G, H, I, J, K).