Teacher Evaluation In Saudi Arabia Relative To National And International Teacher Evaluation Standards And Best Practices

Suhayb Kattan
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/3860
Teacher evaluation systems are essential in tracking the quality and effectiveness of teachers and education systems. This dissertation examines the teacher evaluation system in Saudi Arabia through multiple perspectives, including research-based frameworks for teacher evaluation, norms and best practices from OECD countries, and established Saudi Teaching Standards. An explanatory sequential mixed methods model was used. Data collection involved 643 responses to a teacher survey that covered working conditions, existence of professional learning communities, and teacher evaluation practices, methods, and uses. Following the survey, in-depth interviews were conducted with 11 teachers and 3 key informants.

Findings from this study indicate the evaluation system is not functioning well. Teachers report having lost confidence in its effectiveness and relevance. The evaluation system does incorporate best practices established in other countries. There is an overdependence on principals as sole evaluators. And, because principals are largely untrained for this role and the purpose of evaluation is not clear, the judgements are typically subjective and inconsistent. Teachers report that they seldom receive feedback from their evaluations, and it is not associated with incentives or consequences.

Among recommendations, reform of the teacher evaluation system would benefit from an alignment with objectives of the education system and existing Saudi Teaching Standards. Systemic change is needed, implying that diverse stakeholder groups need to be informed and
engaged in the process of reforming how teacher evaluation is practiced and used in Saudi Arabia. Reflecting on best practices and insights from informants, there is a need for the teacher evaluation system to pursue both formative and summative purposes. Although revising teacher evaluation may need strong central guidance and mandates, it is recommended that this be accompanied with some degree of decentralization so that schools can adapt evaluation strategies which suit their communities; this should also help ensure ownership of the reform process. Among other recommendations, the following should be highlighted: (i) use of evidence-based models, (ii) promotion of professional learning communities to support and improve evaluation practice and use, and (iii) efforts to institutionalize teacher evaluation; in other words, ensuring the involvement of all stakeholders, and integrating evaluation into the daily work of schools.

This study is significant and makes notable contributions. Because of a scarcity of research on the Saudi teacher evaluation system, the detailed description of this system is important for increasing understanding of current practices and shortcomings. The recommendations synthesized from informants and those generated by the researcher provide insights and ideas on how to reform teacher evaluation and better align this with existing teaching standards and best practices established in other countries. It will be difficult to reform the Saudi teacher evaluation system without understanding the intricate and interrelated components and the diverse mechanisms and stakeholders involved. For this reason, the study presents a comprehensive framework for teacher evaluation that can help reform and develop the teacher evaluation system. Finally, the study contributes to the field of evaluation and social science research by providing a sound illustration on how to implement sequential mixed methods designs.
ACKNOWLEDGMENTS

All praises and gratitude are due to Allah, the most gracious and the most merciful. It is He who enabled me to pursue this through all the blessings and bounties He provided. I hope that He accepts this work from me as a little contribution to making the livelihood of teachers and our new generations better.

To my dearest parents, Abdulhakim and Um Kalthoom, my beloved wife, Banan, and my cherished children, Lamar and Sahel. Thank you all for your patience and support throughout this long journey. You have been constantly encouraging and motivating, and without you all, this work would have not been completed.

I would also like to thank my supervisor and dissertation committee chair, Dr. Gary Miron, for his guidance, encouragement, and expertise in ensuring this work would be completed. To my committee member, and qualitative and mixed-methods educator, Dr. June Gothberg, thank you for all the expertise and knowledge you have shared with me and taught me. To Professor Joseph Kretovics, my committee member, thank you for allowing the time and expertise to critically examine my work and help improving it.

Suhayb A. Kattan
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .................................................................................................................. ii

LIST OF TABLES .......................................................................................................................... viii

LIST OF FIGURES ......................................................................................................................... ix

CHAPTER

I. INTRODUCTION .................................................................................................................... 1

   Background of the Problem .................................................................................................... 4

   Statement of the Problem ....................................................................................................... 6

   Purpose Statement .................................................................................................................. 9

   Research Questions ............................................................................................................... 9

   Significance of the Study ...................................................................................................... 10

   The Structure of the Dissertation ........................................................................................ 11

II. LITERATURE REVIEW ......................................................................................................... 12

   Evaluation ............................................................................................................................ 12

      Different Approaches to Evaluation .................................................................................. 14

         Quantitative vs. Qualitative ......................................................................................... 14

         Summative vs. Formative ............................................................................................. 14

         Internal vs. External ....................................................................................................... 16

   Teacher Evaluation .............................................................................................................. 18

   Why Teacher Evaluation? .................................................................................................... 19

   Quality in Education ............................................................................................................ 20
Table of Contents—Continued

CHAPTER

Accountability in Education ................................................................. 20
Teaching Standards and Frameworks ....................................................... 22
The Context of Saudi Arabia ................................................................... 26
The Education System in Saudi Arabia ................................................... 26
The Structure of the Education System .................................................... 30
Teacher Evaluation in Saudi Arabia ......................................................... 30
Saudi Teaching Standards (STS) ............................................................... 31
Teacher Ranking System .......................................................................... 33
Methods of Teacher Evaluation ............................................................... 35
Classroom Observations .......................................................................... 35
Student Surveys ....................................................................................... 37
Student Achievement Scores and Value-Added Models ............................. 38
Teacher Self-Evaluation .......................................................................... 40
Portfolio Assessment ................................................................................ 41
Summary .................................................................................................. 42

III. METHODOLOGY ................................................................................ 43
Study Design ............................................................................................ 49
The Quantitative Methods ...................................................................... 51
The Teacher Survey ................................................................................... 52
The TALIS 2018 Results .......................................................................... 58
Table of Contents—Continued

CHAPTER

The Analysis of Quantitative Data ................................................................. 60
The Qualitative Methods .............................................................................. 60
   Document Analysis .................................................................................. 61
   The Teacher Interviews .......................................................................... 61
   The Key Informant Interviews ............................................................... 63
   The Analysis of Qualitative Data ............................................................. 65
   Trustworthiness of the Qualitative Data .................................................. 66
Limitations .................................................................................................... 67

IV. FINDINGS .................................................................................................. 69
   Document Analysis of the Best Practices of Teaching Standards .......... 70
   Teachers’ Perception of the Saudi Teacher Evaluation System ............ 71
      Teachers’ Working Conditions .............................................................. 73
      Teacher Evaluation ............................................................................... 75
      Teacher Professional Development ..................................................... 89
      Teachers as a Professional Community .............................................. 94
   Comparative Analysis of the Saudi Current Teacher Evaluation and the OECD Best Practices ................................................................. 95
      Existence and Framework of Teacher Evaluation .............................. 95
      Procedures of Teacher Evaluation ...................................................... 96
      Evaluators ............................................................................................ 99
Table of Contents—Continued

CHAPTER

Consequences of Teacher Evaluation ................................................................. 101

The Alignment Between Teacher Evaluation and the Saudi Teaching
Standards (STS) ................................................................................................. 103

Teachers’ and Key Informants’ Perceptions on the Challenges and
Recommendations for Teacher Evaluation ...................................................... 109

Challenges Facing Teacher Evaluation in Saudi Arabia ......................... 110

Recommendations in Accordance with OECD Best Practices ................. 112

Recommendations by Teachers .................................................................... 115

Recommendations by Key Informants ......................................................... 118

Conclusion ..................................................................................................... 120

V. DISCUSSION ............................................................................................... 123

Overview of the Dissertation ....................................................................... 123

Recommendations and Policy Options for Teacher Evaluation in Saudi Arabia .. 127

Teacher Evaluation Policy Options ................................................................. 130

A Hint of Decentralization ............................................................................. 131

Clarity ............................................................................................................. 132

A Developmental Focus ................................................................................ 132

Evaluation for Teacher Benefit ................................................................... 133

Professional Learning Communities ............................................................ 134

Evaluation Consequences ............................................................................ 135

Diverse Trained Evaluators ......................................................................... 137
Table of Contents—Continued

CHAPTER

More Stakeholder Engagement ................................................................. 138
Significance of the Research ................................................................. 139
Recommendations for Future Research .............................................. 140
Conclusion ................................................................................................. 142

REFERENCES ............................................................................................... 144

APPENDICES ............................................................................................... 158

A. The Current Teacher Evaluation Observation Form and Its Translation .......... 158
B. The Saudi Teaching Standards and Its Translation .................................. 161
C. The Teachers’ Survey (English) ............................................................ 164
D. A Cross Walk Table to Align Items From Different Surveys .................. 174
E. Teacher Interview Protocol ................................................................. 180
F. Key Informant Interview Protocol ....................................................... 182
G. Survey Reliability and Factor Analysis .............................................. 185
H. WMU HSIRB Approval ......................................................................... 190
LIST OF TABLES

1. InTASC Model Main Domains................................................................. 24
2. Danielson Framework for Teaching (FFT) Main and Sub-domains ................ 25
3. The Use of Different Methodologies to Answer the Research Questions ........ 51
4. Demographic Groups Percentage Comparison Between the Population and the Different Samples.......................................................... 55
5. Descriptive Statistics for Key Indices.......................................................... 57
6. The Comparison Indicators........................................................................ 60
7. Teachers interviews participants demographics ........................................ 63
8. Descriptive Statistics................................................................................. 72
LIST OF FIGURES

1. The Study Socio-ecological Framework ................................................................. 46
2. The Different Phases of the Research Process ......................................................... 50
3. Conceptual Mapping of Themes in TALIS 2018 ....................................................... 59
4. Themes and Subthemes Layout .............................................................................. 73
5. Teaching Working Conditions as Reported by Teachers ........................................... 75
6. How Many Times Do Formal Evaluations Occur? ..................................................... 76
7. Stakeholder Involvement in Teacher Evaluation ...................................................... 78
8. Methods of Evaluation Used in Evaluating Teachers ............................................... 80
9. Evaluation Feedback ............................................................................................... 81
10. Discussion of the Evaluation Results ..................................................................... 82
11. Level of Awareness of the Teaching Standards ...................................................... 83
12. Awareness Means .................................................................................................. 84
13. Representation of STS in the Current Evaluation .................................................... 86
14. Personal Consequences of Evaluation ................................................................... 87
15. Consequences for Evaluation of Teaching Skills .................................................... 88
16. Evaluation Type .................................................................................................... 89
17. Required Professional Development ....................................................................... 92
18. Types of Professional Development ...................................................................... 93
19. Teacher Engagement in Professional Community Activities ................................. 94
20. Teacher Evaluation Framework .............................................................................. 129
CHAPTER I
INTRODUCTION

The quality of education that children receive is a critical issue in countries around the World. Teachers are a critical component of school quality. A key challenge policymakers and school leaders face is knowing how to identify and develop high-quality teachers. Having a teacher evaluation system in place does not necessarily address this challenge. In many countries with teacher evaluation systems in place for a long time, it is still hard to distinguish between high- and low-quality teachers (Weisberg et al., 2009). Hence, researchers and policymakers around the world are continually trying to figure out improved procedures and methods to increase the precision of education quality measures, including teacher evaluation (Newton et al., 2010).

In the past, researchers believed that schools and teachers did not have much impact on student achievement. Coleman et al. (1966) state that “schools bring little influence to bear on a child's achievement” (p. 325). This idea gradually faded as more sophisticated statistical procedures were used to investigate the complex relationships between school and teacher characteristics and their effect on student achievement. For example, the research done on the Tennessee Value-Added Assessment System (TVAAS) has boosted the importance of the teacher-related factors in shaping student achievement (Sanders & Horn, 1998; Wright et al., 1997). Despite all the discussions about the validity of the TVAAS, it is still one of the most prominent examples of value-added models and its ability to demonstrate a teacher's role in student achievement. A growing body of research on this subject has shown that schools can make a difference when it comes to student achievement (Heyneman & Loxley, 1983).
Specifically, teachers' attributes and qualities are among the most critical factors influencing student achievement (Dodeen et al., 2012; Goe et al., 2008).

Since the late 1960s, the concept of teacher effectiveness has been developing through recognizable phases. One of the earliest phases could be seen in the work of some sociologists such as Coleman et al. (1966) and Jencks et al. (1972). They wanted to investigate the contribution of schools to society at the time. Thus, their focus was on aggregated school performance, which showed a minimal effect of schools on the greater society. Campbell et al. (2012) discussed these skeptical findings. They related the skepticism to several factors, amongst which is the reliance on measures of cognitive gain while undermining other possible measures of school effectiveness. This shows how distinct groups of stakeholders or researchers would define teacher/school effectiveness to what fits their agendas and positions. Nevertheless, all these groups of researchers still agree that teachers are the main factor in any educational system and that they can shape students' academic and non-academic futures (Burroughs et al., 2019; Chetty et al., 2014; Goe, 2007).

Another more developed phase of research on teacher effectiveness is related to the rise of numeration in the educational field during the late 20th and early 21st centuries. Under this surge, teacher effectiveness has been related directly to the ability of a teacher to influence student scores on standardized tests. This is not a surprise, as the whole education system in the West was shifting towards all kinds of standardized tests. The problem with such a view on teacher effectiveness, however, is that it implies causality between how effective a teacher is and student test scores. Such an assumption is an oversimplification of the complex nature of the learning and teaching processes. Many other factors could enormously affect how a student
performs on a test. Reducing all these factors to teacher effectiveness is a gross simplification (Goe et al., 2008).

A more comprehensive view of the construct of teacher effectiveness considers multiple components to the quality of a teacher. Goe et al. (2008) state that five components will lead a teacher to be effective. The first is the teacher's elevated expectations for all students. One way of measuring this is with achievement scores and the exploration of students' perceptions of their teachers' expectations for them. The second component of teacher effectiveness is the active contributions of teachers towards positive academic, attitudinal, and social outcomes of their students (Campbell et al., 2004). Third is the various usage of resources to engage learners. The fourth is the contribution to a diverse classroom and school environments. Lastly is collaboration with other groups within the school community such as other teachers, administration, and parents on the benefit of student success. Hattie (1999) argues that teacher effectiveness is enhanced as a function of the feedback given after evaluation. The comprehensive view of teacher evaluation aligns with the view of Hattie (1999). Exclusive dependence on student achievement to evaluate teacher effectiveness does not provide needed feedback for both stakeholder groups: teachers to enhance their teaching effectiveness, and policymakers to improve education policy. Stated differently, this comprehensive view of teacher effectiveness considers evaluating the inputs, which are put into the system by a teacher, the teaching processes, and finally, the outcome of teaching (i.e., student achievements) (Kane, 2013).

The challenges in reaching a high-quality education system are an international concern. Countries from around the world are coming together through different Intergovernmental Organizations (IGOs) to establish guidelines and frameworks to aid in raising the quality of education. Saudi Arabia, in specific, has started opening up internationally during the last decade
through participating in some International Large-Scale Assessments (ILSAs), looking for ways to strengthen its education system. In 2018, the country participated for the first time in the Organization for Economic and Co-operation Development (OECD) international assessment surveys, the Program for International Student Assessment (PISA), and the Teaching and Learning International Survey (TALIS). The results of these assessments revealed how weak the educational system is. Nevertheless, these assessments have generated data that is of high value and usefulness, as it is the first time to have such databases that are representative of the whole education system. This dissertation contributes to the movement of educational reform in Saudi Arabia through diverse ways. One way is through the acknowledgement of the teachers’ voice about what they think of teacher evaluation current practices. The second way is through use of some international data, best practices in the field, and well-developed teacher evaluation frameworks to find out what is missing and what should be done about it.

**Background of the Problem**

By looking at several data sources and reports about the Saudi education system, one can see the enormous job the government has undertaken in developing an educational system from scratch since the unification of the kingdom in 1932 (Roy, 1992). More details on the development of the Saudi education system will be presented in the next chapter. This development has been partially successful and should never stop, especially when indicators from various sources suggest the existence of some challenges facing the educational system. Even though the government had imposed several national educational reforms, student achievement on some of the ILSAs—such as PISA and the Trends in International Mathematics and Science Study (TIMSS)—are still significantly below the international average, and they are not getting any better.
During the past few years, the Saudi government has planned and undertaken an unprecedented reform, “Saudi Vision 2030,” which covers most aspects of government, including the educational system. This reformative vision is partially viewed as a reaction to some of the flaws seen in the educational system. The main pillars for this vision, as stated by the government (Saudi Vision 2030, n.d.), are: (1) a vibrant society, (2) a thriving economy, and (3) an ambitious nation. Even though advancing the economic status of the country is the main driving force for this vision, the government did not neglect the importance of quality education as a means of empowering the economy. The vision, however, is so generic when referring to advancements in education that it did not point directly to any policies or indicators which will help in guiding the achievement of its ambitious educational goals.

Teacher evaluation in Saudi Arabia has been done since the formal education system was founded in the 1970s. Alqarni (2015) states that there are no other methods of teacher evaluation used in the Saudi public education system other than classroom observations. School principals and educational supervisors are responsible for conducting classroom observations at least once every year to evaluate teachers. The same observation instrument is used for all subjects and grades (Appendix A). The results of these observations are hardly used for formal rewards or sanctions, neither for schools nor for individual teachers. Such an approach to teacher evaluation is insufficient in reflecting the effectiveness of teachers and teaching processes. Darling-Hammond (2012) argues that it is unfair for teachers, students, and the community not to have a standards-based evaluation system for teachers, which incorporates different input methods and, at the same time, promotes professional development among teacher communities. She also states that various sources of data should be integrated into the system to reach a clearer picture of who is contributing to the advancement of students and who needs support and additional
resources. For the ambitious vision of 2030 to become a reality, the reliance solely on subjective evaluations of individual teachers, rather than a multifaceted and more objective evaluation, will not help in improving the quality of education and teachers; hence, educational outcomes will not be any better than the past (Bunaiyan, 2019).

The existence of international data about the working conditions of teachers and on student achievement (e.g., TIMSS and PISA) is one prominent place to start to explore the teacher evaluation system. Comparative studies between the Saudi teacher evaluation and that of other countries will also ease the development process as not to start where others had already started years ago. Besides, the development process should engage some of the instruments applied internationally and check for their appropriateness to be applied in the kingdom. In short, the wheel should not be recreated. Yes, there should be a consideration of the contextual differences between Saudi Arabia and other countries when applying any international instrument; however, the existence of these instruments and methods will serve as a starting point, which will ease and shorten the process.

**Statement of the Problem**

The main problem investigated in this study is the misalignment between the current practices of teacher evaluation in Saudi Arabia and the Saudi Teaching Standards (STS). Such misalignment is contributing to several complications in the educational system. Among the most evident of these complications is the fact that policymakers are not able to assess the soundness of the STS due to the lack of data they have. The effect of the absence of such data will stretch all the way to affecting the educational system outcomes; that is, student achievement. Furthermore, if the teacher evaluation system is relying on one data source, for
which reliability has not even been established or tested, little could be done to improve the effectiveness of teachers and the quality of the educational system.

One primary reason for the current dilemma is the fact that, until recently, there have not been any clear standards for the teaching profession in Saudi Arabia. As a result, teachers have little ability to know the basis on which they are going to be evaluated. Professional development also suffers due to inadequate expectations for teachers. Solving part of the problem, the Education and Training Evaluation Commission (ETEC) has recently published the STS (2017), which should be regarded as the core of the teacher evaluation system. However, the practice has not changed yet, and the old traditional way of evaluating teachers is still in effect, making it hard for the STS to be of any usefulness.

The misalignment between the STS and teacher evaluation practices is also creating a hole in the data needed by researchers and policymakers to act upon. Bridging this gap is significant, as it will allow for deeper understandings of the challenges and issues facing teachers and students in the classroom, and therefore lead to workable solutions. For example, there needs to be a more comprehensive teacher evaluation system in terms of input sources—so not to depend merely on classroom observation done by the principals, but rather a more comprehensive package of methods should be embraced. In addition to leading to lack of data, the misalignment also contributes to some practical challenges in the classroom. For instance, reliable and objective differentiation between high-performing teachers and practices from inferior ones is not possible, incentives and deterrents are not effective, and teachers gradually lose motivation towards actively engaging in their profession. The lack of feedback, corrective recommendations, and targeted professional development will also lead to teachers being unaware of areas they should develop and what are the best practices in the field (Hakim, 2015).
Al Shannag et al. (2013) discussed these issues while comparing Saudi and Singaporean teachers. They state that unlike Singaporean teachers, Saudi teachers tend to overestimate their practices and abilities, which are not reflected in student achievement.

Lastly, the misalignment could have an impact on student achievement levels. In the recently published results of the PISA 2018, Saudi students scored 88 points below average for participating countries in reading, 116 points below average in mathematics, and 103 points below average in science (Echazarra et al., 2019). Saudi Arabia has also been participating in the Trends in International Mathematics and Science Study (TIMSS) since 2003. During all the previous four cycles (i.e., 2003, 2007, 2011, and 2015), Saudi Arabia has shown a deficient achievement level on the TIMSS. This low performance depicts, in a way, the difficulties facing the Saudi educational system. For instance, the average math score for 4th-grade Saudi students on the TIMSS 2015 was 383 points with a standard error of 4.1, and that is significantly lower than the TIMSS scale center-point of 500 (Martin et al., 2016). Additionally, a recent report published by the ETEC (2019) states that more than half of the students in 4th and 8th grades did not meet the minimum standard in both science and math on the first administration of the national tests which took place in 2018.

These pieces of evidence from various sources depict the low achievement of Saudi students and indicate that something taking place in the classrooms is not right (Burroughs et al., 2019; Chetty et al., 2014). Since teachers are such a critical component of schools, one main path to address these challenges includes an effective teacher evaluation system. This system should reflect on real practices, issues, and struggles facing both teachers and students in the classroom and act upon it. A first step would be to include setting standards for the teaching profession so that teachers are fully aware of what is expected from them. These standards, as Darling-
Hammond (2013) stresses, are the only path towards professional development that covers the needs of both, students, and teachers.

ETEC (2017) published the professional paths and the STS; however, the teacher evaluation system in Saudi Arabia has not done much to integrate and consider these STS and paths. This delay has resulted in the STS not being effective in boosting teacher performance. It has also resulted in challenges inside the classroom remaining there, as there are no means to unlock the classroom’s black box. With all the efforts being made toward developing the educational system in Saudi Arabia, it is essential to have a teacher evaluation system that can measure the effectiveness of a teacher in the classroom. Those teachers are the primary mechanism of enforcing any newly developed plans for the education system, so their effectiveness and performance should be evaluated.

**Purpose Statement**

The aims of this dissertation are twofold. First, a descriptive and comparative investigation will be conducted to explore the Saudi teacher evaluation system that is currently in place. That system will be contrasted to other countries’ standards and best practices, and to the Saudi Teaching Standards (STS). This will aid in deepening the understandings of the key features and flaws of the current Saudi teacher evaluation system and help find ways to improve it. Second, based on best international practices, a framework and a set of recommendations will be put forward to ensure better teacher evaluation that is aligned with the STS.

**Research Questions**

The questions guiding this inquiry are as follows:
1. What are the main characteristics of the Saudi Arabian teacher evaluation system, and how does Saudi Arabia differ from the other countries in terms of teaching standards and best practices?

2. How does the current Saudi Arabian teacher evaluation system compare to the Saudi teaching standards?

3. What steps need to be taken to ensure alignment between teacher evaluation practices and the Saudi teaching standards?

**Significance of the Study**

The findings of this study add to the limited body of knowledge about teacher evaluation in Saudi Arabia. They also help educators and researchers in knowing more about how teachers in Saudi Arabia are evaluated in comparison to other countries. The use of international data in comparing the Saudi teacher evaluation system to other countries provides insights about similarities and differences in different teacher evaluation systems. The construct of teacher effectiveness was examined in the Saudi context, and key informants' thoughts about how it is currently measured were collected. These thoughts provide insight for both academics and policymakers. The framework resulting from this investigation also adds to existing teacher evaluation theories. Future researchers may find it useful to test or validate some of the instruments that are part of the framework. The Saudi government may also find the products resulting from this dissertation such as the steps forward to be useful, and possibly adapt and implement them.

The study also contributes to the concept of teacher evaluation. It covers the concept from the perspective of Saudi teachers in a unique fully segregated educational system. The
outputs and findings add to the knowledge about teacher evaluation in general and about the developmental stages of teacher evaluation in a developing country.

The Structure of the Dissertation

The dissertation will follow a five-chapter layout. In the first chapter, the topic of the study is briefly introduced, the problem which the study investigates is presented, and the main research questions and objectives are introduced. The second chapter is comprised of a literature review about the different concepts discussed in the dissertation. Teacher effectiveness related theories are discussed. Evaluation, in general, is introduced with more concentration on teacher evaluation frameworks and standards. The Saudi educational system is also presented with a focus on the teacher evaluation aspect of it.

In the third chapter, the methodology of the study is presented. It also contains all data sources and data collection protocols, procedures, and samples used in the study. Approaches to data analysis are additionally discussed in this chapter. The results of the analysis are presented in the fourth chapter. Finally, in Chapter V, a discussion about the results takes place, with a recommended framework to move forward. A set of recommendations and policy options is, therefore, presented.
CHAPTER II
LITERATURE REVIEW

This chapter introduces the main concepts related to teacher evaluation and how it is addressed in the literature. It will help in framing the conceptual background which this study is based upon. First, the concept of evaluation is introduced, where the main evaluation theories are briefly stated. Different approaches to evaluation are also described. Next, the chapter focuses on personnel and teacher evaluation, in specific. Ideas such as the importance of teacher evaluation, its relation to the concept of accountability, its connectedness to the quality of the educational outcomes, the primary teacher evaluation frameworks, the importance of teaching standards in teacher evaluation, and the difference between teacher quality and teaching quality, are introduced here as well. Then, a more specific discussion takes place about the Saudi education system, and the teacher evaluation policies currently in place. Finally, the different methods of teacher evaluation that are commonly used are introduced.

Evaluation

Evaluation is a practice that has been used for an exceedingly long time. In fact, since the dawn of history, people could decide on what suits them and what is better for their existence, regardless of what better entails. The ability to judge things, people, and ideas is the core concept of evaluation. The broad spectrum it covers has led to having several definitions of what evaluation is. Fitzpatrick et al. (2011) state that there is not yet an agreed-upon definition for the term evaluation. Linguistically, there are about 60 different terms that are related to evaluation, and apply to one context or another (Scriven, 1991). The root term of evaluation is value, so there should be a value, or a set of them, when conducting evaluations. Those act as a reference in the determination of an objects worth or significance. Many evaluation professionals refer to
the basic and simple definition of evaluation that was put forward by the Joint Committee on Standards for Educational Evaluation (1994) which indicates that evaluation is a systematic assessment of the worth or merit of an object. This definition is advantageous due to its conciseness in stating the core of evaluation, and its genericity so that it applies fairly to all different approaches to evaluation. Nonetheless, advocates of different approaches to evaluation will tweak this definition to emphasize whatever approach they introduce.

Another widely used definition is the one by the encyclopedia of evaluation where evaluation is defined as “an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, product, person, policy, proposal, or plan” (Fournier, 2005, p. 140). This definition of evaluation depicts the wide variety of contexts and situations where evaluation may be involved.

It could be argued that the basics of evaluation is that it is a process that involves collecting and synthesizing data and judging them against a set of values to reach a decision (Boulmetis & Dutwin, 2011). The evaluation process may differ from context to another, and that will usually depend on the main objectives of the evaluation. This fact also resulted in having many different approaches to evaluation, which will be covered in the next section.

Examples of these different approaches' definitions may include the approach found by Patton (2008), the utilization focused evaluation, where he emphasizes the aspect of utility and engaging both the intention and the users of the evaluation in how he defines his approach. Another example is Stufflebeam and Coryn (2014), who extended the Joint Committee's 1994 definition and included operational steps and values in the definition. They state that evaluation “… is the systematic process of delineating, obtaining, reporting, and applying descriptive and
judgmental information about some object's merit, worth, probity, feasibility, safety, significance, and/or equity” (p. 696). Bamberger and his colleagues (2012) also established a context of what they call “real world evaluation” where constraints evaluators usually face are described in depth and dealt with. Their definition is more functional than theoretical and aims to provide a decent level of adherence to evaluation standards within real world constraints. The constraints they discussed in their book are time, data, budget, and political constraints.

**Different Approaches to Evaluation**

**Quantitative vs. Qualitative**

It could be argued that in evaluation, as well as in social science research, the golden standard used to stem from the positivist realm. Quantitative methods and experimental designs were leading the field. The work of Donald Campbell (1957), for example, sets a positivist approach towards evaluation. He was considered the “methodologist of the experimenting society” (Shadish et al., 1991). On the other side of the spectrum lies the constructivists. Methods that will dig deep into the lives and experiences of people are needed to investigate complex and socially constructed interventions. In the 1960s, Michael Scriven started what he later relates to as utilization-focused evaluation (UFE) (Scriven, 1966). The focus of the UFE approach is to dig deep into the interactions amongst diverse types and groups of stakeholders to gain a solid knowledge of the social intervention in hand. Bridging the practical, not the theoretical, gap between these two poles, mixed methods research (MMR) is becoming increasingly common to strengthen the flaws of each method; that is, qualitative and quantitative (Ofek, 2016).

**Summative vs. Formative**

Michael Scriven (1966) differentiated between the goals and the role of evaluation. As the originator of the terms *formative* and *summative* (Mertens & Wilson, 2019), Scriven argues...
that the evaluation goals have always revolved around the idea of giving merit, worth, and value to the evaluand (i.e., the evaluated entity). However, the role of evaluation may be highly variable. Patton (2008) states that determining merit and worth are not necessarily judgmental but could serve an improvement goal. It may be part of an audit for an educational program, an activity for supporting underperforming teachers, or an activity of data gathering about a recently developed curriculum. The evaluation goals and how the evaluation results will be used are to determine the significant role of the evaluation. If, for example, the main intention was to eliminate poorly functioning objects then an evaluation is summative, whereas if the intention were to assess and improve the utility of the evaluand then the evaluation is formative. Mertens and Wilson (2019) differentiates between formative and summative evaluations by looking at the time the evaluations took place. If the evaluation was happening during the process of a program, for example, then it is considered formative whereas if it happened at the end of a process, a program for example, then the evaluation is summative.

Ford and Hewitt (2020) built a conceptual framework for teacher evaluation that encompasses both objectives, the summative or as they put it “organizational interest,” and the formative “the personal interest.” They argue that the need of such framework is inevitable for having a teacher evaluation system that does not trigger teacher stress and maintains a prominent level of teacher satisfaction. In contrast, Popham (1988) argues that combining the two types of teacher evaluation (i.e., formative and summative) in one system renders both dysfunctional. He argues that the results from each type of evaluation is different; hence, they cannot be combined and one of them should be the primary motive for the system. The design of the evaluation and all the utilized tools will be affected by such decision. Wholey (1996) takes a middle approach arguing that for policymakers it is the formative aspect of evaluation that matters the most. They
need information and data to assist them in decision making. Moreover, he argues that the summative part of the evaluation is also important to control quality of processes. He recommends finding opportunities to “fine tune” performance evaluation systems aid in continuous decision making throughout the year with an end of year summative evaluation.

**Internal vs. External**

Technically, internal evaluations are the types of evaluations that are done internally within an organization. They are conducted through internal evaluators who are already part of the organization. Internal evaluation is mostly related to monitoring and quality control. Hence, they tend to be more formative than summative. On the other hand, external evaluations are done by external evaluators specifically hired to do the evaluation, and they primarily serve the accountability aspect of evaluation (Scriven, 1991; Vanhoof & Petegem, 2007).

In the context of schools, scholars have emphasized the need to navigate away from evaluations that are solely external (Christie et al., 2004; Nevo, 1994; Vanhoof & Petegem, 2007). The collaboration of both types of evaluations, internal and external, could be done in different ways as suggested by Christie and her colleagues (2004). For example, an internal evaluator could be responsible for collecting the data and the external evaluator would be responsible for analyzing the data. Another way of approaching this collaboration is conducting two separate evaluations and then discussing them collectively to compare and integrate the results. It all depends on the purpose behind overtaking two types of evaluation. The baseline is that combining internal and external evaluations will increase the benefits of both and will enhance the autonomy of the school (Nevo, 1994).

Darling-Hammond (2012) discussed the idea of Teacher Professional Learning Communities (TPLCs) as being an alternative to internal evaluations in schools. She advocates
for creating such collective communities where teachers have the chance to distribute the
knowledge and skills they retain, and they also can provide the support needed by some teachers.
There is not a clear definition on what TPLCs are exactly (Prenger et al., 2019; Stoll et al., 2006;
Vangrieken et al., 2017). However, the main characteristics of such groups that are usually
brought up while discussing them include, autonomy, the large impact veteran teachers have in
them, the collaboration between teachers, the growth promoting environment, and the operation
as a collective enterprise (Vescio et al., 2008). Stoll et al. (2006) argues that the word
“communities” in TPLC is key to the concept. They state that after spending decades trying to
improve school systems through focusing on individual teachers, it is time to redirect the
attention to community development, as it is proven to be much more effective and sustainable.

Researchers during the past decade tried to investigate the effectiveness of TPLCs. For
instance, the meta-analysis done by Vescio et al. (2008) explored several studies that
investigated the effect of such communities on teacher and student learning. Collaboration
among teachers is higher when working in TPLCs. New and developed instructional strategies
are also observed in schools that implements TPLCs. Above all the results comes the
contribution of TPLCs in promoting continuous teacher learning. It also helped in facilitating the
efficiency of professional development. They end by introducing several studies that show how
student learning has increased due to TPLCs.

To sum up, scholars have presented solid evidence that TPLCs could be affective in
promoting a healthy sustainable loop of development and improvement through collecting data,
analysis, reflection, and change. Teacher evaluations will then be a natural process done in an
informative, internal, supportive manner with a sense of belonging and being accountable (Stoll
et al., 2006).
Teacher Evaluation

Evaluating the performance and effectiveness of people differs enormously from evaluating any other entity. People have feelings and they respond to evaluators, which makes the process of assessing their performance complicated. On one hand, evaluation decisions might have a direct harsh effect on a person and may put him and others dependent on him in a harsh situation if, for instance, he was expelled from his job. In this regard, scholars also stress that personnel evaluation systems are prone to being harmful to people, so it is of a great importance that these systems are occasionally subjected to critical and systematic reviews to ensure fairness and justice (Kaufman et al., 2006).

On the other hand, the complex and vital role teachers play in an education system makes it necessary to evaluate their performance and eliminate any underperforming teachers from the system (Kellaghan & Stufflebeam, 2003). In addition to that, the complex role of teaching adds to the challenges of getting teacher evaluation right. What teaching should look like? What are the criteria for quality teachers? How can these criteria be measured reliably?

Dwyer and Stufflebeam (1996) pictured the necessity and complexity of teacher evaluation beautifully by stating,

Teacher evaluation is a highly controversial area, with myriad stakeholders and a wealth of technical, psychological, political, ethical, and educational complexities. Teacher evaluation is relevant to every segment of the educational system, and society at large has an intense interest in how it is carried out and what its impact on education and on individuals’ lives will be. Thus, the criticisms of theory and practice are strongly held, and how (or whether) these criticisms are resolved has direct implications for the quality of American schooling. (Dwyer & Stufflebeam, 1996, p. 768)

More than 20 years later, teacher evaluation remains a controversial topic and things seems to have not progressed dramatically.
Stronge (2006) states two main elements of teacher evaluation. These elements align to a great extent with Green's (2006) painless performance evaluation approach. The first element is about recording the level of performance of a teacher, and second is to support teachers to develop their performance. By professionally applying both elements, an education system ensures that the accountability aspect of teacher evaluation is justified. He also states several attributes of a teacher evaluation system that makes it effective. The first attribute is having mutually beneficial goals for both the system and the individual teacher so both the summative and the formative goals of teacher evaluation can be achieved. The second attribute is effective and positive communication between the teachers and the evaluators as representatives of the educational authority. Such communication helps in the inclusion of teachers as main players in the planning and execution of the evaluation plans. The third attribute of an effective teacher evaluation system is that it should create a climate for a quality evaluation where the trust and the cooperation between the teachers and the evaluators are not just encouraged but systematized in some form. The fourth attribute is a technically sound evaluation system, which makes it easier for stakeholders to participate in the system and take it seriously. Finally, the last attribute of an effective teacher evaluation system is the utilization of multiple data sources to increase the validity and reliability of any evaluation decisions. Methodologies of evaluating teachers tend to lean towards being subjective, hence the importance of adding multiple data sources to insure reaching higher accuracy in the evaluation results.

**Why Teacher Evaluation?**

Teachers in many countries have been treated as scapegoats soon after the publication of large-scale international testing results (Papanastasiou, 1999). However, research has proven that the complexity of an educational system is unique. The number of human factors (i.e., teachers,
administrators, students, and parents) in the reception of a series of inputs and signals will surely provide outputs that vary significantly, no matter how much control an educational system has. In most cases, the argument behind using teachers as scapegoats lies upon the fact that teachers are the most influential when it comes to student success. As the next sections will show, this fact is backed by scientific research that shows how influential teachers are in shaping their students’ futures (Al Shannag et al., 2013; Chetty et al., 2014; Rowan et al., 2002; Shuls, 2018; Wilson & Wood, 1996). This has led Papanastasiou (1999) to argue that it is crucial to have an evaluation system that ensures an acceptable level of teacher quality and help them build up their skills and ability to teach and create a difference.

**Quality in Education**

Cheng and Tam (1997) state that the term *quality* in the management literature has no agreed upon definition. Instead, the term may be used to refer to value, utility, conformance to standards or specifications, or the ability to avoid flaws. They continue to argue that in education, things are also vague and disputable. Educational system quality is looked at from different perspectives, the inputs, the processes, and the outputs and outcomes.

**Accountability in Education**

Accountability in education has garnered an increasing amount of focus over the past couple of decades. One main motivation for this increased focused is the effect of international organizations (e.g., OECD), which, in one way or another, are putting much pressure on countries through the international assessments they regularly conduct (Garver, 2019). In the center of educational accountability lies teachers, who are historically described as the “most responsible and notable members of society because their professional efforts affect the fate of the earth” (Helen Caldicott, 2008 as cited in Larsen, 2010; Skedsmo, 2019). In the Islamic
tradition as well, teachers are referred to in the Holy Scriptures as being the most valuable and in the highest status of significance right after God’s prophets. This central position of teachers has led communities and governments to hold teachers accountable (Alhamid et al., 2007).

Teachers being the center axis of any educational system causes them to be under severe pressure from both governments and societies (Connell, 2013). This pressure is mainly seeking answers to crucial questions about public education. Are teachers satisfying what we, as a country, expect from them? Furthermore, do the educational outcomes meet the specified standards? Since the establishment of modern schooling systems, there have been several stages of schooling and teacher accountability. Webb (2005) identifies several bases for educational accountability. First is the purpose of educational accountability. Researchers in democratic nations have been struggling between what Webb (2015) refers to as professional vs bureaucratic accountability. The essence of this struggle is that professional accountability, as advocated for by many (Kostogriz & Doecke, 2011; Lissovoy & McLaren, 2003; Verger et al., 2019), tends to put more power in the hands of the professional which, in many cases will result in them being inclined to their interests. On the other hand, a bureaucratic accountability, done by the state for mainly for summative purposes, will fall short in recognizing the real struggle, work, and concerns teachers have. However, the focus of accountability for the past decade has resulted in an increase attention on standardized tests. In many countries, the only form of school and teacher accountability is done through high-stake testing standards-based reforms. Kostogriz and Doecke (2011) argue that the current focus on standardized testing and standards-based reforms is only serving the bureaucratic regimes in posing more control on what is taught and how it is taught. They admit that evidence is still scarce on the effect of such standard-based accountability and reform on the learning outcomes and the quality and effectiveness of teachers.
Second in relation to the bases of accountability in education is the effect of accountability on the educational system. Is the accountability system in place playing a role of building capacity and professionalism or rather focusing on how to punish and reward teachers and schools (Kostogriz & Doecke, 2011)? Holloway (2020) discusses the term datafication of the accountability systems in education. She argues that the current surge in using and mandating high-stake testing and standard based improvements is transforming the attention from the real practice and struggles to the numbers on datasets. Teachers’ expertise, authority, and accountability will consequently plunge, she argues. The use and importance of data could not be denied in the current era; however, as this “datafication” is transforming education, a line should be drawn somewhere to stop surveilling teachers and students as this will drastically destroy the capacity and professionalism of the teaching force.

A balance between the two main types of accountabilities, professional and bureaucratic, is desirable but hardly attainable (Firestone & Donaldson, 2019). Kostogriz and Doecke (2011) proposes what they call the profissional ethics, which consitutes the empowerment of teachers from within the teachers’ community to admire their job and their profissional responsibility towards the future of their students. They argue that such approach is going to fuel teachers to do their best and regardless of the circumenstances they face everyday.

**Teaching Standards and Frameworks**

At the beginning of the 20th century, teaching was considered a profession of those who cannot do better, as George Bernard Shaw once said (Shaw, 1903). In Western countries, teachers are also the first to get blamed for economic, social, and educational issues and problems. The movement towards teaching standards started in the West during the mid-20th century (Call, 2018). It aimed at the time to transform teaching into a profession (Cochran-Smith
& Zeichner, 2005). Until now, there is a paucity of scientific evidence that shows the effectiveness of such standards; however, they are used in many education systems as a quality assurance maintaining strategy (Paraschivescu, 2017).

In the United States, the McKinsey and Company report (How the World's Best-Performing School Systems Come out on Top; Barber & Mourshed, 2007) was an eye-opener for the fact that “the quality of an education system cannot exceed the quality of its teachers” (p. 16), and that the primary route towards a better education system is through raising the status of teachers and improving teaching quality. The National Board for Professional Teaching Standards (NBPTS), created in 1987, aimed to establish “rigorous and high standards of what accomplished teachers should know and be able to do” (Darling-Hammond, 2013, p. 21). These standards, as Darling-Hammond (2013) states, have informed several essential developments in the American education system since the 1990s, amongst which is the various aspects of the teacher licensure program. The standards also aided in reconceiving teacher assessment. It allowed evaluators to be more specific and seemingly objective by assessing attributes that are highly influential in making a teacher more effective. The specifics of these standards also made teachers’ professional development more efficient and directed towards the essential needs of a teacher. It is a step in the direction of professional accountability, where teachers gain autonomy and power from inside of their profession (Webb, 2002); however, Darling-Hammond (2013) stresses the idea that these standards are not a magic wand that will fix issues related to teacher effectiveness by just having them. Instead, she explains how, in many states, the existence of these standards had no value added to the teaching profession. The amount of joint work that should be put in place alongside the standards is crucial for effective implementation. In the subsequent sections, some of the available teaching standards will be explored.
The NBPTS (n.d.) proposed five core propositions for teaching standards that are to be the basics of any teaching standards. Those are:

1. Teachers are committed to students and their learning.
2. Teachers know the subject they teach and how to teach those subjects to students.
3. Teachers are responsible for managing and monitoring student learning.
4. Teachers think systematically about their practice and learn from experience.
5. Teachers are members of learning communities.

The Interstate Teacher Assessment and Support Consortium (InTASC), which is part of the Council of Chief State School Officers (CCSSO), has also contributed to the body of knowledge about teacher effectiveness and the ways needed to support ongoing teachers. The CCSSO (2013) outlines the standards that are necessary to be demonstrated by new teachers. The InTASC has ten standards grouped into four main categories, so they are easier to think about and follow (see Table 1).

Table 1

*IntASC Model Main Domains*

<table>
<thead>
<tr>
<th>Category</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>The learner and learning</td>
<td>1- Learner development</td>
</tr>
<tr>
<td></td>
<td>2- Learning differences</td>
</tr>
<tr>
<td></td>
<td>3- Learning environments</td>
</tr>
<tr>
<td>Content</td>
<td>4- Content knowledge</td>
</tr>
<tr>
<td></td>
<td>5- Application of content</td>
</tr>
<tr>
<td>Instructional practice</td>
<td>6- Assessment</td>
</tr>
<tr>
<td></td>
<td>7- Planning of instruction</td>
</tr>
<tr>
<td></td>
<td>8- Instructional strategies</td>
</tr>
<tr>
<td>Professional responsibility</td>
<td>9- Professional learning and ethical practice</td>
</tr>
<tr>
<td></td>
<td>10- Leadership and collaboration</td>
</tr>
</tbody>
</table>
Last in this brief introduction to the primary teaching standards is the Danielson Framework for Teaching (FFT). In the 2013 version of the FFT (Danielson, 2014), which is the latest, Danielson states that the FFT aims to categorize the empirically proven aspects of teaching that will increase the effectiveness of an individual teacher. It defines what teachers should know and do to increase their students' learning. It also aims to align between the new Common Core State Standards (CCSS) and the teacher evaluation standards. The philosophy underlying the FFT is centered on keeping students intellectually active while in the classroom. Table 2 presents the main and sub-domains of the FFT. Under each sub-domain, the FFT gives a brief introduction to its core concept and its elements. It also lays out some of the indicators that could be used to assess the fulfillment of the sub-domain and the levels of proficiency.

Table 2

Danielson Framework for Teaching (FFT) Main and Sub-domains

<table>
<thead>
<tr>
<th>Domains</th>
<th>Sub-domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning and Preparation</td>
<td>a. Demonstrating knowledge of content and pedagogy</td>
</tr>
<tr>
<td></td>
<td>b. Demonstrating knowledge of students</td>
</tr>
<tr>
<td></td>
<td>c. Setting instructional outcomes</td>
</tr>
<tr>
<td></td>
<td>d. Demonstrating knowledge of resources</td>
</tr>
<tr>
<td></td>
<td>e. Designing coherent instruction</td>
</tr>
<tr>
<td></td>
<td>a. Designing students’ assessment</td>
</tr>
<tr>
<td>The classroom environment</td>
<td>a. Creating an environment of respect and rapport</td>
</tr>
<tr>
<td></td>
<td>b. Establishing a culture of learning</td>
</tr>
<tr>
<td></td>
<td>c. Managing classroom procedures</td>
</tr>
<tr>
<td></td>
<td>d. Managing students’ behavior</td>
</tr>
<tr>
<td></td>
<td>e. Organizing physical space</td>
</tr>
<tr>
<td>Instruction</td>
<td>a. Communication with students</td>
</tr>
<tr>
<td></td>
<td>b. Using questioning and discussion techniques</td>
</tr>
<tr>
<td></td>
<td>c. Engaging students in learning</td>
</tr>
<tr>
<td></td>
<td>d. Using assessment in instruction</td>
</tr>
<tr>
<td></td>
<td>e. Demonstrating flexibility and responsiveness</td>
</tr>
</tbody>
</table>
Table 2—Continued

<table>
<thead>
<tr>
<th>Domains</th>
<th>Sub-domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional responsibilities</td>
<td>a. Reflecting on teaching</td>
</tr>
<tr>
<td></td>
<td>b. Maintaining accurate records</td>
</tr>
<tr>
<td></td>
<td>c. Communicating with families</td>
</tr>
<tr>
<td></td>
<td>d. Participating in the professional community</td>
</tr>
<tr>
<td></td>
<td>e. Growing and developing professionally</td>
</tr>
<tr>
<td></td>
<td>f. Showing professionalism</td>
</tr>
</tbody>
</table>

The Context of Saudi Arabia

The Education System in Saudi Arabia

Hakeem (2012) states that education in the Arabian Peninsula has emerged since the existence of Islam in the early 7th century. Mecca and Medina, the two holy cities of Islam, were educational hubs for people from around the region. The main focus of education at that time was religion (Alkhedr, 2011). This remained the same for centuries. Parents will usually send their children to the Ktateeb, which is a traditional way of schooling that is usually located in mosques (Davidson, 2008). Students in these Ktateeb would be expected to memorize the Holy Quran and get the basics of the Islamic religion and teachings. Basic math and the Arabic language were taught as well. Teachers at the Ktateeb were usually religious scholars who have full authority to discipline the students and teach them whatever they think is essential knowledge.

In the early 19th century, there were few modern schools in the area of Hijaz, the western region of the Arabian Peninsula, where the holy cities of Mecca and Medina are located. The Ottoman Empire controlled these few schools. Turkish was the primary language for teaching in these schools; hence, families were hesitant to send their children to these schools (Hakeem, 2012).

The modern Saudi education system was established in 1953 when the Ministry of Education (MoE) was founded. The mission of the new ministry was to build a centralized
education system from scratch. The government demand was high compared to the socio-economic nature of the country at the time. A massive number of citizens were not yet settled in cities; rather, Bedouins traced the rain in search of lands suitable for pasture. It explains the nation’s high illiteracy rate of about 90 percent in the initial stages of the new country. Most people were against the modern schooling system, partly because they were rigidly attached to traditional schooling (i.e., the Ktateeb), which was perceived as a better option for them culturally and religiously. This resulted in opposition to the modern schooling system (Alsalloum, 1991).

In 1970, the MoE issued the official education policy and system document. This document highlights the basic principles of the new education system. Unsurprisingly, the document derives most of its articles and clauses from the Islamic heritage and culture, which was the main foundation of the newly developed country. Alminqash (2006) describes how this policy document defines the philosophical framework of the education system, its goals, its types, and its stages. She continues to state the main attributes of the education system, in light of the education policy document, which sets it apart from other education systems around the world. These include the fact that it was fundamentally built upon Islamic theology and principles, the right for all citizens to receive free education, and the complete segregation between males and females in all educational stages. Although it has been about 50 years since publishing this policy, there has not been any revisions or developments in it. As Alminqash (2006) states, the policy was appropriate in meeting the societal and economic needs at that time, but it has not been modified since then. Al-Essa (2009) stresses how critical it is to reorganize the priorities of the education system and focus on the essential needs of the 21st century, as those are barely evident in the current policy document (Al Shahwan, 2012).
This “outdated” policy document, as described by Rugh (2002), is a portrayal of the broader political posture in the kingdom. The Saudi government is regarded as a monarchy where the royal family is in control of all the regulatory aspects of the country, including education. The King, who is the prime minister as well, appoints all ministers in the council of ministers counting the minister of education. The standard length for a minister to serve in a position is 4 years, and it may be extended or shortened by the King’s order. Rugh (2002) points out that such a custom represents, in a way, the authoritarian role that the government has on the education system. Curricula in all public and private schools are enforced by the government. Besides, teacher recruitment, evaluation, and sanctions are all administered by the government (Al-Essa, 2009). It is thus clear that the Saudi education system has a “centralization” endowment that it cannot discard.

Forty-six education directorates spread across the country and connected directly to the MoE in Riyadh (Ministry of Education, n.d.). Each directorate oversees several superintendence offices, which superintend several schools in its area. This hierarchical system has the MoE at its top with complete control over all the inputs and processes of the whole education system. In general, having such an authoritative system confines the leadership skills of those on the bottom of the hierarchy, and it restricts the creativity of all personnel within the establishment (Hui & Bonan, 1991).

Despite all the domestic and international critiques of Saudi education, the MoE was successful in achieving some accomplishments since it was founded. For instance, the literacy rate had increased to about 70 percent by the early 1990s and 95 percent in 2017 (UNESCO, n.d.). In that same year, the number of schools reached a peak of more than 36,000 schools according to the education and training survey that was done in 2017 by the General Authority.
for Statistics (GAS). The same survey also reported that the net enrollment rate in primary education had reached 98 percent, while the gross enrollment rate is 100 percent. These figures reported by the GAS match what is reported by The World Bank (n.d.).

In addition to these accomplishments, the Saudi government has been spending generously on education during the last two decades. The average government expenditure on education is 20 percent of the total government expenditure, which reflects the importance of education to the Saudi government (Khatib, 2011; “The Saudi budget, education, health, and services comes first,” 2019). Many reforms took place over the years, trying to tackle the issue of education quality (Al-Essa, 2009; Allmnakrah & Evers, 2020; Alyami, 2014; Elyas & Al-Ghamdi, 2018; Tayan, 2017). Nevertheless, the public is still not so convinced that these reforms are doing what they are intended to do. There seems to be something lacking that is causing all these efforts to be ineffective.

Allmnakrah and Evers (2020) have presented facts about Tatweer, a formidable education overhaul that began in 2007. The Saudi government allocated more than three billion U.S. dollars for this project. They state that even after 10 years of the implementation of Tatweer, it has failed to create its aimed reform. Al-Essa (2009), who later became the Minister of Education and then the Chairman of the ETEC, states that the lack of political will and vision is behind this failure. Alyami (2014) has also investigated the Tatweer project. She listed several accomplishments that are fuzzy and unsupported. For example, she states that the project shifted the focus of the education system from the individual to the collective benefit and advancement. This, however, contradicts the Tatweer project's central policy, where the focus is instead on the individual as Allmnakrah and Evers (2020) state. A significant criticism of Tatweer, which Alyami (2014) points to, is that students’ achievement on the international assessments such as
TIMSS has just gotten worse. She concludes by recommending giving school principals more independence to make decisions and to engage teachers in the thinking process of any reform. By doing so, the MoE would obtain the agreement of teachers, who are the actual executors of any reform.

**The Structure of the Education System**

Compulsory education in Saudi Arabia consists of three main stages: 6 years of primary school, 3 years of intermediate school, and 3 years of secondary school (Saudi Arabian Cultural Mission, 2006). Boys and girls are segregated throughout the educational system with male teachers teaching in boys’ schools and female teachers teaching in girls’ schools.

**Teacher Evaluation in Saudi Arabia**

Alqarni (2015) pointed out that there is a paucity in research about teacher evaluation in Saudi Arabia. It is also noted by Alharthi (2017) that teacher evaluation policies are not easily accessible by researchers, which is a common struggle for researchers in the country. To overcome this lack of information, the primary source that will be used in this study is several dissertation studies that researched teacher evaluation in Saudi Arabia. The survey that will be part of the current research will help gather information about the current practices and policies.

In the traditional practice of schooling (Ktateeb), teachers were in a high-status level, as they are the ones responsible for teaching the new generation and passing the traditions and teachings of culture and Islam to them. Teachers at the time were not promoted to teach until they had shown an elevated level of knowledge, manners, and ability to instruct students. Usually, they are promoted to be taught by their teachers. No authority was to question teachers or to hold them accountable in today’s terms. After the establishment of the modern education system, teachers’ honorable status remained intact for a fair amount of time (Hakeem, 2012).
By examining the education policy document that was issued in 1970 and which outlines the bases of the education system in the country, we can see that teacher evaluation was not mentioned in the entire document. The only mention of teachers was in the context of teacher training and preparation programs. Al Shahwan (2012) investigated the perception of 58 superintendents from around Saudi Arabia. One of his findings was the lack of specified roles for teachers, making evaluating their effectiveness unfeasible. It is hence not so surprising that the Saudi teachers are perhaps at the receiving end of any educational reform that has been implemented thus far. Notwithstanding, teacher evaluation does exist in schools; yet there is a lack of a coherent policy stating the rationale and philosophy behind such policies and its consequences.

The version of teacher evaluation that the MoE in Saudi Arabia still follows is a traditional style, where classroom visits are merely considered. The responsibility of conducting these visits is split between school principals and superintendents. They evaluate teachers at least once a year though in many cases, evaluations do not occur. School and superintendence administrators will often notify teachers before the evaluation visit. Evaluators use a standard observation form that the MoE requires them to use. I have not seen any study based on data that was generated using this form. The form tends to measure teachers’ performance, characteristics, and communication skills. As Alharthi (2017) states, the evaluation form that is used is summative in nature and teachers usually do not anticipate getting any feedback, nor even a chance to discuss their evaluations.

**Saudi Teaching Standards (STS)**

A milestone in the evolution of teacher evaluation in Saudi Arabia is the establishment of the Commission for Public Education Evaluation in 2013, later named the Education and
Training Evaluation Commission (ETEC). In its founding regulation, the ETEC is responsible for evaluating school performance, setting standards for the teaching profession, and building and administering all measures of teachers’ proficiency and the teacher licensure program (National Center for Archives and Records, 2019). Since its establishment, one of the first goals for the ETEC was to build a certification test for new graduates who plan to join the teacher workforce. Dimitrov and Alsadaawi (2014) state that the teacher test aims to provide the MoE with a reliable measure to rely upon in choosing capable and proficient candidates. The teachers’ test measures four domains: (1) professional knowledge, (2) promoting learning, (3) supporting learning, and (4) professional responsibilities.

In addition to the teacher test, the ETEC worked on building standards for the teaching profession. This is one of the most important steps for having an effective evaluation system for teachers (Darling-Hammond, 2012). According to the ETEC (2017), the STS and professional paths align with the VISION 2030, which focuses on enhancing the country’s economy through building a generation that is equipped by the essential skills needed to be creative and innovative as stated in its document. The construction of these standards started by looking at the standards that are applied in other countries. Next, a group of local researchers and experts built the theoretical framework for the standards. Then, several consulting committees, which included educationalists, superintendents, and teachers from different regions of the country, carried out a series of workshops and meetings and put together a draft of the standards. After that, an international educational firm in addition to local and international experts reviewed the standards until they reached a satisfactory level. The standards document also states that ETEC had conducted a public survey about these standards. More than 25,000 individuals participated
in the survey, including teachers, students, educationalists, and parents. Nonetheless, none of the survey results were reported.

The STS regards teachers as the primary element in the education system and states that these standards set the responsibilities, knowledge, and practices which teachers should resemble, know, and master. It also enables teachers to perform their job adequately and effectively, knowing precisely what is expected from them as teachers. The document also adopts a learner-centered teaching approach in which all the instruction processes should have the learner’s needs as its primary focus. These standards should be a reference to the teacher licensure program, teaching proficiency tests, teacher professional evaluation, teacher preparation programs, teacher professional development, teacher professional classification and ranking, self and institutional evaluation, and teacher innovation and research programs.

The STS has three main domains: (1) professional ethics and responsibilities, (2) professional knowledge, and (3) professional practice. There are several primary standards under each domain, and under each primary standard, there are several secondary standards. The standards document also includes a rubric that precisely describes the level of proficiency expected form teachers on each secondary standard and for each rank (i.e., practitioner, advanced, and expert). Appendix B presents the standards in Arabic and its translation to English.

**Teacher Ranking System**

The professional teachers’ ranking system has three different ranks. The basic is the practitioner teacher, divided into two primary levels (probation and practitioner). A teacher in probation is a novice who just joined the profession after completing one of the certified teacher preparation programs. They also should meet the requirements for the interim teaching license.
Those teachers will be put under probation for the first 2 years, where the focus is to extend their skills, knowledge, and experiences under the supervision of the superintendent, the principal, and other expert teachers. Within these 2 probation years, a novice teacher should be able to demonstrate an intermediate level of proficiency in the administration of the standards at the practitioner rank. If so, they would be promoted to being a practitioner teacher. A practitioner teacher does meet the satisfactory level of the professional teaching standards; thus, they would get their permanent teacher license. Teachers at this rank are expected to collaborate effectively with other teachers at their schools and extend their instruction skills and knowledge. They are also expected to demonstrate skills and knowledge expertise on the theories and practices in relation to the teaching standards at the practitioner level.

The second rank is the advanced rank. Teachers at this rank should be competent in the teaching standards of this rank and should meet the requirements for its teaching license. They should show commitment to advance their skills and knowledge base. They should also collaborate with other teachers and exchange their experiences to raise the instruction level at their school. The last rank is the expert teacher rank. Teachers are promoted to this rank after demonstrating a proficiency level in related standards and meeting the requirements for its teaching license. At this rank, teachers are expected to be “researcher teachers” in the sense that they can solve the problems they face in their classes using a data-based method. They should also lead the support of less-experienced teachers at their schools by running workshops and training sessions. They are also expected to engage in formative evaluations of their colleague teachers and encourage them to be creative in their teaching.

From looking at these rankings and their requirements, one can interpret that the STS and ranking system are setting a new way of looking at teacher effectiveness in Saudi Arabia. It
requires that teacher evaluation be done consistently and effectively for the ranking system to work as aimed. The focus on collaboration between teachers and the support they are expected to show to each other elevates the status of the teacher evaluation system in the country. However, so far, there are no practical methodology guides nor instrumentations for teacher evaluation.

**Methods of Teacher Evaluation**

In this section, different methods of teacher evaluation will be introduced. The purpose of this section is to identify different methods of teacher evaluation presented and discussed in the literature. Arguments for and against each method will be presented. Amongst these methods are classroom observations, lesson studies, peer collaboration, student ratings, student achievement, portfolio, and self-evaluations.

**Classroom Observations**

Many educational systems around the world rely heavily, and sometimes solely, on classroom observations to determine the effectiveness of teachers (de Lima & Silva, 2018; van der Lans et al., 2016). Several evaluation models also include classroom observation as a method of data collection about teacher effectiveness. For example, the Danielson Framework (Danielson, 2014) proposes a classroom observation tool that is used in different states around the United States, and that has been investigated by several researchers for its validity and reliability.

Classroom observations are used to measure the instructional skills and performance of teachers in addition to their ability to maintain a healthy learning environment in the classroom (Montgomery, 2002). The main cause for including such method in teacher evaluation is the logic that teachers’ practice should improve through providing feedback to the teachers after classroom observations take place. It is also thought that these observations aid in linking
teachers’ practice to the teaching standards that are in place, which are believed to increase teaching effectiveness and students’ achievement to the desired levels. Above all these motivators to use classroom observations to evaluate teachers comes the feasibility aspect. It is one of the most affordable and straightforward methods to reach a decision about a teacher proficiency level, but of course with so many drawbacks. Amongst the crucial criticism of classroom observations comes its lack of reliability and validity (Lei et al., 2017; Polikoff, 2015; Steinberg & Garrett, 2016). Steinberg and Donaldson (2016) relates this to the fact that many of the observation tools used in classroom observations are not grounded in theory or research, rather they are, in most cases, done by the effort of administrators at the school or district levels. In addition, relying on one classroom observation to judge teachers is a widespread practice.

Such practice, obviously, may be substantially biased due to many external factors (van der Lans et al., 2016). van Der Lans (2016) and his colleagues have studied the reliability of this method. They conclude that to reach a reliability of 0.70, which may be sufficient within a formative framework, a minimum of three observation visits are required. However, if the classroom observations are done within a summative framework, the acceptable reliability should be no less than 0.90, and that requires more than 10 observation visits. It is that practice that also resulted in not having enough validation studies for these types of observations. These findings align with what Hill et al. (2012) and Sandilos et al. (2019) have found.

While applying classroom observations as a tool for teacher evaluation policymakers and evaluators should be cautious about some aspects of the evaluation to ensure a relatively higher accuracy in the evaluation results (Bergin et al., 2017). de Lima and Silva (2018) state that training evaluators is key to a valid application of classroom observations. What is usually done by most education systems is that school principals are responsible for conducting the classroom
observations. Those principals are already having trouble with the overwhelming number of tasks they are responsible for, let alone the lack of training they get in terms of observing teacher performance in classrooms. Politics within the school community will also affect the evaluation results, according to de Lima and Silva (2018).

**Student Surveys**

Students are the core and primary beneficiaries of any education system. Thus, many researchers advocate for the perceptions of students to be part of teacher evaluation systems. Sandilos and her colleagues (2019) state that the information produced by students about the instructional and pedagogical practices in the classroom is valuable. It is still an under-developed method especially at the initial stages of education; however, if carefully built and administered, these surveys will better our understandings of teacher effectiveness. Advocates of such method usually discuss the use of such methods for a formative constructive aim rather a summative one. However, on the other side of the argument there are several scholars who criticize such approach, as it is prone to be biased. In fact, there is very little done about the validity of students’ ratings in public education compared to what is done on the topic at the tertiary education level (Downer et al., 2015; Kuhfeld, 2017; Marsh et al., 2019; Sandilos et al., 2019).

The Tripod Survey is a well-recognized survey for measuring students’ perceptions on several domains related to teacher effectiveness (Tripod Education Partners, 2016). It is aimed to provide information to educators about the instructional quality and the classroom management through what they call the 7Cs—care, control, clarify, challenge, captivate, confer, and consolidate (Tripod Education Partners, 2016). Different studies have validated the factorial structure of the Tripod Survey, and all show moderate to high reliability. However, the factor structure is still not coherent at this point (Kuhfeld, 2017).
Sandilos and her colleagues (2019) investigated the convergent and predictive validity of three different methods of teacher evaluation, amongst them was the Tripod survey. The overall conclusion they reached is that at the secondary level there where stronger evidence that these different methods are measuring the same construct (i.e., teacher effectiveness). However, at the elementary level the convergent validity was less evident. Such results are suggestive that the use of students’ surveys in teacher evaluations are more valid at higher levels of public education. However, it might be an instrument weakness that caused the results not to be as valid in the elementary level (Ferguson, 2011, 2012).

**Student Achievement Scores and Value-Added Models**

Educators and evaluators have used students’ scores on standardized tests as a straightforward method to decide on a teachers' effectiveness for a long time. In the early 1990s, value-added models (VAM) started spreading in the United States to utilize the pool of students’ test scores on standardized tests in the process of evaluating teachers. VAMs are algorithms that utilize students’ test scores each year as their input and then adjust them according to several covariates to get the teachers’ added value every year.

Several studies have found that teachers count as a primary variable in explaining the variations between students’ scores (Nye et al., 2004; Rockoff, 2004). Some went even further by arguing that the value-added for specific teachers is highly predictive of the value-added for the same teachers in the following years (Gordon et al., 2006). Equipped with the results from such studies accompanied by the governmental push in what is known as the Race to the Top (RTT) policy, the number of states applying VAMs as ways to hold teachers accountable has increased significantly (Hill et al., 2011). However, the process of determining the effectiveness of a teacher is far more complicated and troublesome than just calculating value-added scores.
The complex nature of teaching and learning, as well as the multiple roles a teacher is expected to perform in the classroom, makes the reliance on VAMs in evaluating teachers an undermining procedure. The validity of VAM scores has been under scrutiny by researchers for the last 2 decades. One of the main controversies surrounding their use is that research is not firm on the specifications of these models. Some argue for the need to include students’ and schools’ covariates to adjust for students’ differences, while others would argue that the usage of value-added scores from previous years is sufficient in accounting for students’ differences, as these scores will be used as a baseline in the model. This criticism arises from the fear that the sorting of students to teachers does not account for the use of these models in evaluating the teachers (Chetty et al., 2011). A more technical criticism for using VAMs is that they merely rely on students’ scores, which have a ceiling effect (Hanushek & Rivkin, 2010). This will underestimate the variance components of teachers’ effects on students, which will ruin the mere concept of trying to measure the effect of teachers on student learning. Individual teachers’ value-added scores will be highly influenced by the underestimation of the variance components, and hence the reliance on such a method to evaluate teacher performance would be problematic.

It is also worth noting that such models should not be applied without considering their limitations and shortcomings. Corcoran (2010) lists several limitations to the use of VAMs that should be understood by educators and policymakers before any of these models is put to work. Primarily, what does the student achievement score used in the VAM measure exactly? This limitation opens a huge window to criticize the use of such models, as pointed to by Baker et al. (2010). They argue that not all subjects taught to students could be measured annually for a specific grade content. It is hard to major the annual change in many subjects, especially the ones related to humanities, as these are not tied to specific age group or grade level. Another
limitation is the appropriateness of the measurement tool. Is there enough evidence to prove the validity and reliability of the scores produced by these measures? It is also crucial to be aware that any change in the VAMs could be attributed to changes in the school as a whole and not just a factor reflecting teacher effectiveness.

One more limitation to applying VAMs is its dependance on the availability of students' standardized test scores of multiple years. In Saudi Arabia, for instance, there is no national standardized testing system that covers all school grades; hence, VAMs would not be a feasible option for evaluating teachers in the country. That is also why the researcher was not able to locate any studies about the topic in the context of Saudi Arabia.

In summary, the main methods used in teacher evaluation are described above. What to use from these methods is a decision that should consider many varied factors. A crucial guideline is that no one method should be solely utilized. Muñoz and Dossett (2016) emphasize the importance of having more than one method of data collection when it comes to evaluating teachers. This is one way to ensure we are stepping towards more valid, just, and meaningful teacher evaluations (Sandilos et al., 2019).

**Teacher Self-Evaluation**

Teacher self-evaluation (reflection) has emerged to enhance participatory methods of evaluation where the teacher (the evaluand) is actively participating in the process (Wear & Harris, 1994). The definition of a reflective teacher has been discussed in the teacher education literature and mainly includes the idea of learning from experience and being aware of areas of practice that are in need for attention and development (Taylor, 1994). Taylor argues that an essential part of self-evaluation is that teachers are more likely to engage and act upon data that they themselves have participated in gathering. For them, the validity of the data they provide is
much more reliable and valid than those gathered by principals or superintendents. Teacher autonomy is another related concept that is usually brought to the discussion about teacher self-evaluation (Cirkovic-Miladinovic & Dimitrijevic, 2020). The contribution a teacher adds to their own professional development through self-reflection and self-evaluation aids in achieving the evaluation aims with less pain and apposition. However, Towndrow and Tan (2009) argue that power is inevitable in any evaluation system, as the criteria are already set by the authority and all that teachers are doing through the practice of self-evaluation is to provide data about how close they abide by the rules.

As a proponent of self-evaluation, Taylor (1994) argues that self-evaluation is valuable regardless of the systematic actions that might result because of it. She also states that it would be so useful even if the only result of such practice is to raise teachers' awareness of their practice and teaching. However, she believes that in most cases there are far more long-term benefits than one can imagine.

**Portfolio Assessment**

Teaching is a complex process that is hard to measure. In most cases, it is hard to put our hands on tangible evidence that teaching and learning have happened. The idea of creating portfolios started at the student levels to enable educators to assess the outcomes of their teaching and to provide support to students according to their individual needs (Wolf, 1996). Likewise for teachers, portfolios address the problem of shortage in evidence of their work by providing them with a systematic way to document all the work and effort they are practicing (Bullard, 1998). It is also considered a way of self-evaluation, where a teacher documents the progress and the challenges they face in the classroom and discusses it with their peers or supervisors (Taylor, 1994). It is, however, a time-consuming activity if done professionally,
which is why a systematic unified template is encouraged. This not only makes it easier for the teacher to document, but also makes the process of evaluating the portfolio more smooth and just (Painter, 2001). Painter (2001) also stresses the fact that these portfolios should not be a simple collection of materials, but rather they should be built in accordance with the teaching standards and teachers should be able to justify the inclusion of any document in their portfolio.

Summary

The literature review was intended to lay the foundations of teacher evaluation and to discuss what scholars in the field have found to be the best practices. The chapter provided an intro to what evaluation is based on its intent. Whether the aim is to provide formative feedback to teachers, to grade their performance, or both, will have to be thoroughly thought out, as it will affect the procedure of the evaluation. An introduction to personnel evaluation and its principles was also introduced in this chapter. The constructs of education quality and teacher quality were discussed. The chapter also introduced the reader to the Saudi education system. It is apparent that evaluating teachers on the job is a complicated task that should be carefully planned. There is a plethora of factors that affect the evaluation and its outcomes. Breaking down these factors into manageable categories help to develop a teacher evaluation system that works in specific contexts. It is also crucial to recognize the complex relationships network of teacher evaluation. The next chapter introduces the methodology of this dissertation.
CHAPTER III
METHODOLOGY

This chapter introduces the research design and methodology of this study. The design used in this study enabled the researcher to describe the current teacher evaluation system in Saudi Arabia and to compare it to current practices in other countries. It also aided in producing a teacher evaluation framework and a set of recommendations and policy options for the teacher evaluation system in the country.

Research designs are considered crucial in an empirical investigation. Whether described explicitly or implicitly, research designs remain that piece of the research process that will guide the researcher’s thinking and help them answer the questions they are proposing. The design acts as a guide map through several decisions and procedures during the investigation, starting with the broad, abstract ideas and thinking, moving to data collection and analysis, and ending with extracting meaningful results and conclusions. The design researchers choose for their research journey could be viewed as a thread bringing all the pieces of jewelry together, forming a harmonic piece of literature. Social science researchers approach their investigations using different methodologies (Creswell & Creswell, 2018). The primary three research methodologies are qualitative, quantitative, and mixed methods research (MMR).

MMR is in the middle of the continuum between quantitative and qualitative approaches (Creswell & Creswell, 2018). It mainly involves collecting quantitative and qualitative data during the research project's span to search for more in-depth and additional insights that are not reachable by using one research approach (Poth, 2018). Like many aspects of MMR, its definition is still not agreed upon in the community (Plano Clark & Ivankova, 2016). Different definitions focus on specific aspects of the research process, one of which the researcher who is
giving the definition regards as the most critical aspect of MMR. For example, many definitions focus on the method part of the research process, accounting for the basic fact that MMR must combine quantitative and qualitative data. Of course, that is the core of MMR. However, there are more aspects of MMR than just the mixing of methods. The philosophical combination and the integration process are examples of aspects of the MMR process beyond the mere combination of qualitative and quantitative data.

In terms of the philosophical basis, the quantitative approach is mostly based on a positivist philosophy, whereas the qualitative approach is more into the interpretivism and constructivism philosophy realms. However, for the MMR approach, the philosophical basis is still under debate (Creswell & Plano Clark, 2007; Plano Clark & Ivankova, 2016). Researchers have debated MMR’s philosophical basis and whether it has its distinguished philosophical paradigm or mixes the quantitative and qualitative paradigms (Ghiara, 2019). Ghiara argues that both standpoints are adequate to support the MMR approach and that the controversy stems from how one defines the word paradigm. What is relevant here is that when researchers view paradigms as a representation of different worldviews (such as pragmatism, constructivism, positivism, etc.), then MMR does mix and combine these worldviews in one study. According to Mertens (2012):

The pragmatic paradigm has been put forth as a philosophical framework that supports the use of mixed methods based on the assumption that there is not one set of methods that is appropriate; rather, the criteria for choosing methods include the following: What fits with the research question in this study? (p. 256)

The pragmatist approach is also useful in dealing with MMR’s main feature: the relatively high level of complexity in problems MMR investigates. Mertens (2012) refers to these social and educational highly contested problems as “wicked problems” due to their overly complex nature and the intricated study designs they require to address all their interactions. Poth (2018)
builds her work in MMR on this concept of complexity. She defines it as “characterizing the behavior of a research system whereby its components (such as research participants, researchers, their environments) interact in multiple, nonlinear ways without direction” (Poth, 2018, p. 5). Plano Clark and Ivankova (2016) refer to the components of the “wicked problem” as involving “multiple domains and contexts” (p. 14). In her discussion about complexity, Poth (2018) argues that all research is complex, and that complexity should be regarded as a continuum rather than being a fixed characterization. She differentiates between three primary levels of complexity: low, moderate, and high. Her criteria for this categorization are based on the level of stability of the research assumptions; to what extent the evidence of influence is predictable; the definability of the integration need, steps, and the study boundaries; the level of agreement on the integration amongst the researchers; and the level of predictability of some outcomes of the integration process. As discussed below, the problem this dissertation is trying to investigate seems to lend itself to the moderate level category on the Poth’s complexity continuum. The use of a pragmatic paradigm for this MMR may be justified by how complex the concept of teacher evaluation is. Other philosophies and research traditions would not have the capability MMR has to disentangle the complex nature of teacher evaluation.

Plano Clark and Ivankova (2016) state that “Mixed methods research is a dynamic and interactive process that involves multiple domains and contexts” (p. 14). The domains and contexts they are expressing here range from the personal background and the theoretical and philosophical assumptions to the socio-economic context in which the researcher lives and practices research. These authors also adopt a socio-ecological framework for MMR to guide the interrelationships between all its different components. Figure 1 presents the socio-ecological framework for the current dissertation. The five inner circles represent MMR's main aspects: its
definition, rationale, design, relationship to other approaches, and quality. The three bigger surrounding circles represent the different contexts that influence MMR's application: the personal, the interpersonal, and the social. The different components of the socio-ecological framework will be discussed next.

Figure 1

The Study Socio-ecological Framework
In this dissertation, the main aims are to describe the status of teacher evaluation in Saudi Arabia, to compare it to current practices in other countries, and to build a framework that will align with the STS that were published by the Education and Training Evaluation Commission (ETEC) in 2017. The complexity of teacher evaluation stems from the considerable number of stakeholders who influence its policies and practicalities on the one hand, and on the other hand, the considerable number of stakeholders who are influenced by its consequences and results. Many interactions happen between these stakeholders, making it a “wicked problem” that needs to be investigated using more than one method. As previously discussed, MMR has more aspects of the research process to be blended than the mere mix of methods. Thus, Greene’s (2007) definition of MMR, which illustrates this complexity, will be adopted. She states that “mixed methods social inquiry involves a plurality of philosophical paradigms, theoretical assumptions, methodological traditions, data gathering and analysis techniques, and personalized understandings and value commitments” (Green, 2007, p. 13). The current discussion and explanation of this study’s socio-ecological framework try to explain some of these aspects.

As explained below, the study design serves the complementarity function in that the description of the teacher evaluation system through the survey results is enriched and clarified through the administration of interviews with teachers and key informants. Also, the combination of quantitative and qualitative methods allows for a more rigorous and valid analysis of the practice of teacher evaluation and the teacher evaluation framework that is recommended.

This study’s complex purpose required collecting data from multiple sources and overviewing the entire process through more than one philosophical lens. On the one hand, the study’s descriptive part is more about digging into the real world and trying to find more about the reality of the current teacher evaluation system. An ontologically independent reality of the
teacher evaluation system does exist. The researcher’s role is to collect evidence that will get the community of educators and researchers closer to this reality. On the other hand, building a framework for teacher evaluation is regarded as a methodological contribution to the body of literature. It is as well a practical contribution to the field of education. These contributions will help achieve the community’s greater good, which is manifested in a better teacher evaluation system. Although teacher evaluation does not have one correct or right way to be done, contributions from different researchers will lead to having a better way to evaluate teachers. The philosophical complexity described above is better addressed using an MMR design wearing a pragmatist lens.

In MMR, extra attention should be given to ensuring the quality of the process. A logical extension to the debate on MMR’s philosophical bases is a debate about assessing the validity and quality of an MMR. The literature presents an enormous number of ways to assess MMR’s quality (Dellinger & Leech, 2007; Heyvaert et al., 2013). Plano Clark and Ivankova (2016) recommend three steps to address the quality of an MMR. First, assess the quality of the individual qualitative and quantitative strands of the study. The research would follow the guidelines already in place on ensuring the quality of a qualitative or a quantitative study.

Second, researchers must evaluate the quality of the generated inferences from integrating data from both strands. Quality inferences are regarded as the main deliverable of an MMR and should be of high quality (Tashakkori & Teddlie, 2010). Terms such as inference quality, inference transferability, and legitimation are usually mentioned when discussing the validity of MMR inferences. Teddlie and Tashakkori (2003) first introduced the term of inference quality. It refers to assessing the procedure the researcher uses to draw the research conclusions from the results of the different strands of an MMR. The same scholars introduced the term inference
transferability in 2009. By this, they refer to what is usually known as generalizability in other research traditions. In MMR, transferability describes the ability to generalize the MMR conclusions and apply them to similar contexts. Onwuegbuzie and Burke Johnson (2006) discussed the concept of validity in MMR and produced a typology of legitimation that they argue is an on-going process rather than a fixed attribute of a study. This relates to the continuity and iterative features of MMR, which were discussed earlier.

Lastly, the quality issues of specific MMR designs should be considered. Researchers should present valid answers to questions such as why they chose to investigate the problem using the MMR tradition, why they decided on a specific sequence, and how the design maintains fidelity.

In this dissertation, the primary approach to ensuring high-quality processes included following quality checks for each strand, presenting an argument for the chosen design, and keeping the reader informed of all the integration processes the researcher is implementing. For the quantitative strand, the researcher peer-reviewed the survey to check for any faulty or out of context items. Cognitive interviews with a small number of the population were conducted to ensure all items were well received and made sense to the respondents (content validity), and measures of reliability were reported. In the qualitative part of the study, trustworthiness was addressed through adopting a reflexive lens, the triangulation of both strands of the study, a sample that reflects the demographic variability of the population, and member checking.

**Study Design**

This dissertation followed a sequential mixed method design. Figure 2 shows the overall research process and sequence of this dissertation. During the first phase, the researcher administered a survey that aimed to diagnose the current situation of teacher evaluation in the
country. The initial findings from the quantitative data analysis then informed the process of building the teachers’ interview protocol, which occurred in the second phase of the study. In that phase, teachers were asked to elaborate and confirm some of the survey interesting findings. The initial findings from both, the first and the second, phases informed the key informants’ interview protocol which took place in the third phase. Key informants’ interviews were focused on the policy aspect of teacher evaluation rather than its practice. The researcher tried to solicit information, from key informants, about policy changes and in-progress plans to develop teacher evaluation in the country.

**Figure 2**

*The Different Phases of the Research Process*

![Diagram showing the different phases of the research process.](image)

The three research questions were answered throughout this process while utilizing different methods and analyses. Table 3 shows the type of data that the study used to answer each of the proposed questions. The first question aimed mainly to characterize teacher evaluation in Saudi Arabia and to compare it with some of the best international practices in the field. The teacher survey data, in addition to several documents that represent some best
international practices, were analyzed for that purpose. The second question was concerned with the national context of teacher evaluation. The teacher survey data and teacher interviews were collectively analyzed and integrated to get to the alignment status between the field practice of teacher evaluation and the STS. The third and last question aimed to provide needed steps to align the practice of teacher evaluation to the STS. Interviews of both, teachers, and key informants, aided in providing the basis for these needed steps. The following sections will elaborate on the sampling, procedures, and analysis plan for each of the study phases.

Table 3

*The Use of Different Methodologies to Answer the Research Questions*

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Survey Data</td>
<td>Document Analysis Teacher Interviews Key Informant Interviews</td>
</tr>
<tr>
<td>What are the main characteristics of the Saudi Arabian teacher evaluation system, and how does Saudi Arabia differ from the other countries in terms of standards and best practices?</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>How does the current Saudi Arabian teacher evaluation system compare to the Saudi professional teacher standards?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>What are the steps which need to be taken to ensure a sound alignment between teacher evaluation and the Saudi professional teacher standards?</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

The Quantitative Methods

The quantitative strand of this study aimed to provide sufficient information about the current practices of teacher evaluation in Saudi Arabia to initially enable the researcher to
compare it to some international well recognized practices. It also aided in clarifying any misalignment between the current practice and the STS. To reach these goals, two main quantitative data pools were utilized. The first was the teacher survey, which was constructed by the researcher, as described below. The second source was the summary and results reports issued by the OECD for the TALIS 2018.

**The Teacher Survey**

The survey intended to gather information from a representative sample of current Saudi teachers in public education about the teacher evaluation processes and procedures in their schools. The survey consisted of five main sections, which are: demographics, working conditions, teacher evaluation practices, professional development, and professional learning communities. There were 34 questions on the survey with a total number of 93 items. The English version of the survey can be found in (Appendix C).¹

The demographic section included nine items. Those items include questions about gender, age, years of teaching experience, number of schools the teacher worked at, the primary subject taught, the school level taught, whether the teacher took the teacher admission test or not, and their score on the test if they took it. The second section had four questions about teaching working conditions. The main concentration here is on the time aspect of working conditions, as it positively relates to the teacher’s ability to teach effectively (Hirsch et al., 2006). The third section is considered the core of the survey. It had 12 questions with a total of 50 items. The primary intent here was to gather information about current practices of teacher evaluation in Saudi Arabia. It started by asking about the frequency of the evaluation. It also asked about the nature of teacher evaluation, whether it is formal or informal, and formative or summative.

¹ The Arabic version of the survey is available upon request from the author sak.3231@gmail.com.
Methods, procedures, and consequences of teacher evaluation were also covered in this section. The fourth section included six questions, with a total number of 13 items about professional development. The fifth section had one question with nine items that asked about the existence of professional community support among teachers in the same school.

Existing teacher surveys such as the PISA teacher survey, the TALIS questionnaire, and the Teaching, Empowering, Leading, and Learning questionnaire (TELL) highly informed the construction of the current dissertation's survey. It is worth noting that to attain compatibility of teacher evaluation in Saudi Arabia to different countries, the researcher has put considerable effort into ensuring the survey items align with the TALIS and PISA teacher survey items. A crosswalk table was created to align these three surveys (Appendix D).

**Sample.** The sample for the teacher survey aimed to represent the population of teachers in Saudi Public schools, which, according to the official public Saudi data website (Saudi Data, n.d.), was a little over half a million in 2018. The researcher was in contact with officials from the Ministry of Education (MoE) requesting a sampling framework and communication means with teachers. A negative response was received. However, the contact personal offered distributing the survey for all teachers in the country through the general administration of education in each of the 13 provinces of the country. It could be thought that this method of sampling would result in a more of a simple random sample. However, the response rate for this sampling method was not as anticipated. After all the data cleaning, the number of participants who participated through this sampling technique was 170. The distribution of this sample amongst gender and province is shown in Table 4. From the table, we can see that the cooperation in filling the survey varied vastly. It is obvious that some of the provinces have
either not received the survey at all or that there were no participants willing to take part in the study.

To acquire more participants, a second convenience sample was obtained through distribution of the same survey on social media platforms specialized for teachers. The main platforms used were closed groups on the Telegram application, which had about 50,000 participants, most of them teachers. To ensure participants did not fill the survey more than once, the survey settings were set to deny multiple access to the survey from the same IP address. The number of participants through this sampling strategy was 472. From Table 4, one can see that the percentage of participants in the demographic groups changed, and this sampling method was able to reach to teachers in some of the provinces the first sampling strategy was not able to reach.

By merging both samples, the total number of participants in the survey reached 642. The overall sample distributions amongst the different demographic groups based on gender and province are displayed in Table 4. The acquired sample is convenience rather than random. The sample representation of the different subgroups does not match how these subgroups are distributed in the population. The sample also may be biased towards teachers who are active on social media platforms as it was the main mean of data collection. These and other concerns about the sample of the study will be discussed in the limitations section.
Table 4

Demographic Groups Percentage Comparison Between the Population and the Different Samples

<table>
<thead>
<tr>
<th>Province</th>
<th>Population</th>
<th>First sampling technique (Ministry)</th>
<th>Second sampling technique (Social)</th>
<th>Overall sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Riyadh</td>
<td>43.1%</td>
<td>56.9%</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Makkah</td>
<td>46.3%</td>
<td>53.7%</td>
<td>25%</td>
<td>75%</td>
</tr>
<tr>
<td>Eastern</td>
<td>47.8%</td>
<td>52.2%</td>
<td>52.6%</td>
<td>47.4%</td>
</tr>
<tr>
<td>Asir</td>
<td>46.4%</td>
<td>53.6%</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Madinah</td>
<td>45.6%</td>
<td>54.4%</td>
<td>33.3%</td>
<td>66.7%</td>
</tr>
<tr>
<td>Qassim</td>
<td>43.5%</td>
<td>56.5%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jizan</td>
<td>47.8%</td>
<td>52.2%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tabuk</td>
<td>45.9%</td>
<td>54.1%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ha’il</td>
<td>50.3%</td>
<td>49.7%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jawf</td>
<td>47.7%</td>
<td>52.3%</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Bahah</td>
<td>46.5%</td>
<td>53.5%</td>
<td>11.1%</td>
<td>88.9%</td>
</tr>
<tr>
<td>Najran</td>
<td>44.5%</td>
<td>55.5%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Northern</td>
<td>50.6%</td>
<td>49.4%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>46.0%</td>
<td>54.0%</td>
<td>34.1%</td>
<td>56.9%</td>
</tr>
</tbody>
</table>
**Procedures.** Before the administering the survey, the researcher conducted three cognitive interviews to check for the survey’s content and face validity. Participants in these interviews are current Saudi teachers. During these interviews, the researcher read the survey questions vocally and asked the interviewees to reflect on what they understood from the questions and if it does make sense to them. The input and feedback from these interviews helped in strengthening the survey language and make it more sound to teachers during the real administration. The researcher then submitted an approval application to the institutional review board at Western Michigan University to get their approval for collecting the data. After meeting all required approvals (Appendix H), a pilot administration was conducted to check for any flaws or problems in the instrument in the online platform. The pilot administration had 18 participants. They were asked to provide feedback and assess the wording, and flow of the survey questions. More feedback poured in after this pilot administration, so some wordings were improved. Qualtrics, an internet-based surveying platform, was utilized for the process of data collection. One of Qualtrics’ useful features is that all the data collected is stored behind the Western Michigan security firewall, which adds to securing the data collected from participants. No one other than the researcher had access to the data throughout the dissertation process.

The researcher sent an introductory email (Appendix C) to all participants. The email included a brief description of the survey with a clear affirmation of the voluntary aspect of the participation in the study, the estimated time the participant needs to complete the survey, and a link to the survey page. The first page on the survey contained the consent form and all participants were asked to electronically consent by choosing whether they “accept” or “do not accept” taking part in the study before being able to start the actual survey or directed to an exit page.
Indices. Four indices were created to be used in comparing subgroups for any significant differences. These indices were all calculated by averaging the Likert scale items for each respondent. These indices that were created are: Working Conditions, Standards Inclusion, Administrative Change, and Skills Change. Table 5 presents the descriptive statistics and reliability coefficients for the four indices.

Table 5

Descriptive Statistics for Key Indices

<table>
<thead>
<tr>
<th></th>
<th>No. of items</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SD</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working Conditions</td>
<td>642</td>
<td>7</td>
<td>1.00</td>
<td>4.00</td>
<td>2.18</td>
<td>0.61</td>
</tr>
<tr>
<td>Standards Inclusion</td>
<td>634</td>
<td>15</td>
<td>1.00</td>
<td>4.00</td>
<td>3.36</td>
<td>0.61</td>
</tr>
<tr>
<td>Administrative Change</td>
<td>642</td>
<td>7</td>
<td>1.00</td>
<td>4.00</td>
<td>1.64</td>
<td>0.79</td>
</tr>
<tr>
<td>Skills Change</td>
<td>642</td>
<td>10</td>
<td>1.00</td>
<td>4.00</td>
<td>2.43</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Note. (1 = Totally Disagree, 4 = Totally Agree)

Survey validation. They reliability of the survey was checked for each of the five constructs measured (i.e., Working Conditions, Standards Inclusion, Methods of Evaluation, Administrative Change, and Skills Change), and the whole survey as well. All these constructs were highly reliable with Cronbach Alpha coefficients between (0.82 and 0.95). The overall Cronbach Alpha for the survey was found to be 0.95. For all the constructs, except Standards Inclusion, the reliability was expected to be high, as they are already validated on their original surveys.

An exploratory factor analysis was conducted to check for the survey internal consistency. The results show that there are five main factors representing the Likert type scale
items on the survey (Appendix G). This result was expected, as the survey was initially built to measure five main constructs. On each construct, the items show a moderate to high factor loading on their related factor.

**The TALIS 2018 Results**

The TALIS 2018 cycle included 47 countries and economies with more than 15,000 schools and 260,000 teachers. Saudi Arabia’s data was extracted from the publicly available data files on the OECD website. According to the technical report, the TALIS 2018 followed a stratified random sample procedure for each participating country (OECD, 2019a). For Saudi Arabia, there were 195 participating schools with a total of 2,744 teachers. The stratification criteria were gender (male/female schools) and region (13 different regions). It was in 2018 when the MoE in Saudi Arabia started its participation in the TALIS. This means that the data available from this first participation is still new, which will add to the significance of the current study. Additionally, it would have been of great usefulness if Saudi Arabia had participated in the PISA 2018 teacher survey; however, that did not happen. In the results section, Saudi participation in the TALIS 2018 will be briefly described with the aid of tables and figures.

The main goal of TALIS is to “generate internationally comparable information relevant to developing and implementing policies focused on teachers and teaching with an emphasis on those aspects that affect student learning” (OECD, 2019b, p. 21). Figure 3 introduces the conceptual map of the TALIS 2018. The central theme of TALIS, which aligns with what the current study is trying to investigate, is teacher feedback and development. In this theme, as describe by Ainley and Carstens (2018), the main interest is to look at how teacher feedback and appraisal processes support and affect teacher effectiveness. They also link the concepts of accountability and appraisals of teachers’ professional development effectiveness. Based on
these similarities and the crosswalk table (Appendix D), the researcher specifies a set of indicators in Table 6 that guide the comparison part of the analysis. These indicators cover two main categories: teacher evaluation and professional development.

**Figure 3**

*Conceptual Mapping of Themes in TALIS 2018*

![Conceptual Mapping of Themes in TALIS 2018](image)

The Analysis of Quantitative Data

For the quantitative strand of the study, the researcher conducted a descriptive analysis of the survey data. As this strand is considered the first in the sequential MMR design of this study, the initial findings of the analysis informed the construction of the interview protocol which was used to gather information from teachers. A comparative analysis of the Saudi Teacher evaluation system to the TALIS 2018 study’s main findings was also conducted. This comparison focuses on the constructs where both surveys align using the indicators listed in Table 6. The descriptive and the comparative analyses both aid in answering the first research question of this investigation.

The Qualitative Methods

For the qualitative strand of this study, document analysis will be conducted in addition to two sequentially administered sets of semi-structured interviews. These methods provide the study with thick description and deep understandings of the current situation of teacher
evaluation in Saudi Arabia and how it compares to the international best practices in the field. The qualitative analysis aids in answering all three research questions at various levels.

**Document Analysis**

The main aim for the document analysis is to shed light on some well recognized international practices in the field of teacher evaluation and teaching standards. These practices include the InTASC model and the Danielson framework for teacher evaluation. The TALIS framework and some of its results are also be analyzed. This analysis helps in comparing the Saudi teacher evaluation system to these international frameworks. The results of the document analysis are part of the answer to the first research question.

**The Teacher Interviews**

The first set of interviews in this dissertation targeted teachers who responded to the survey. These interviews were semi-structured and were informed by the initial survey results and aimed to deeply explore the teachers’ thoughts about some interesting findings from the survey. The interviews helped acquiring higher accuracy of the reality of teacher evaluation in the country and how it aligns with the STS.

**Sample.** At the end of the survey, a question was asked to all respondents about whether they would like to participate in an hour-long interview about the survey topic. There were 44 teachers in total who agreed to take part in the interviews. The demographic data of those individuals was revised to make sure that the selected sample represents diverse groups of teachers. The final interviewee sample included 11 teachers, six females and five males. Eight different regions of the country were represented in this sample. Table 7 shows the demographics of the interviewed teachers. The distribution of gender looked more better in the teacher interview sample compared to the survey sample. The distribution od province in the interviews
sample, however, was not as diverse as in it was in the survey sample due to the limited number of interviews the research was able to conduct.

**Procedure.** After reaching the initial results from the first phase of the study, the researcher built the teachers’ interview protocol. An approval request application was then sent to the institutional review board at Western Michigan University to get their approval on conducting the interviews with the teachers. The interview protocol (Appendix E) covers the topics listed below:

- The existence of teacher evaluation
- The frequency of teacher evaluation
- Who is responsible for conducting the evaluations?
- The procedure of the evaluation
- How the evaluations affect teachers
- To what extent the evaluation aligns with the STS

All interviews were conducted in Arabic, as it is the formal language in Saudi Arabia. The interviews were conducted on Zoom, the virtual meeting platform. Because emails are not widely used by teachers, the researcher had to contact the participants through WhatsApp. Each of the participants received an invitation message that had a short thank you message for initially accepting to be part of the interviews, a brief introduction to the topic as they may have forgotten about it, a link to the consent form that could be electronically signed, and a link to the Calendly app, where they could choose a convenient time for them to conduct the interview. All the interviews were audio recorded and transcribed using Sonix, a specialized transcription software.

To protect the confidentiality of the respondents, all personal information, identifiers, and location names were removed from the data transcripts and each interviewee was given a
pseudonym. The researcher then uploaded all the transcripts to the MAXQDA software program to analyze them. The initial results from these teacher interviews and the teacher survey were used to inform the construction of the key informant interview protocol.

**Table 7**

*Teachers interviews participants demographics*

<table>
<thead>
<tr>
<th>Pseudonym</th>
<th>Gender</th>
<th>Province</th>
<th>School level</th>
<th>Age</th>
<th>Subject taught</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ali</td>
<td>Male</td>
<td>Eastern</td>
<td>Middle</td>
<td>33</td>
<td>Arabic Language</td>
</tr>
<tr>
<td>Arwa</td>
<td>Female</td>
<td>Makkah</td>
<td>High</td>
<td>42</td>
<td>Mathematics</td>
</tr>
<tr>
<td>Fahad</td>
<td>Male</td>
<td>Jizan</td>
<td>High</td>
<td>37</td>
<td>Arts</td>
</tr>
<tr>
<td>Hamad</td>
<td>Male</td>
<td>Riyadh</td>
<td>High</td>
<td>38</td>
<td>Arabic Language</td>
</tr>
<tr>
<td>Hind</td>
<td>Female</td>
<td>Qassim</td>
<td>Elementary</td>
<td>27</td>
<td>Science</td>
</tr>
<tr>
<td>Khalid</td>
<td>Male</td>
<td>Makkah</td>
<td>Middle</td>
<td>34</td>
<td>Mathematics</td>
</tr>
<tr>
<td>Nora</td>
<td>Female</td>
<td>Ha'il</td>
<td>Middle</td>
<td>40</td>
<td>Religion</td>
</tr>
<tr>
<td>Rahaf</td>
<td>Female</td>
<td>Riyadh</td>
<td>Elementary</td>
<td>28</td>
<td>Social studies</td>
</tr>
<tr>
<td>Rana</td>
<td>Female</td>
<td>Eastern</td>
<td>High</td>
<td>45</td>
<td>Social studies</td>
</tr>
<tr>
<td>Saad</td>
<td>Male</td>
<td>Madinah</td>
<td>High</td>
<td>39</td>
<td>Arabic Language</td>
</tr>
<tr>
<td>Sara</td>
<td>Female</td>
<td>Tabuk</td>
<td>Middle</td>
<td>46</td>
<td>Religion</td>
</tr>
</tbody>
</table>

**The Key Informant Interviews**

The second set of interviews aimed at expanding the understandings of the topic through the engagement of key informants who are responsible, in some way, for teacher evaluation policies in the country. These were semi-structured interviews as well and focused on significant issues about teacher evaluations and how evaluation is officially viewed. The alignment of
teacher evaluation to the STS and the future of teacher evaluation were also covered in the interviews.

**Sample.** The sampling of the key informants went through some challenges. Initially, the target was to get five key informants who currently work in the MoE or the ETEC and have direct contact or influence upon teacher evaluation policies. Personal connections and communication were the only way to reach out to those people. A list of 10 key informants was developed, and an introductory message was sent to them through WhatsApp. The message included the name of the referring person so that they know how I obtained their personal number, a brief introduction to the research topic, a link to the consent form that could be electronically signed, and a link to Calendly app so they can directly choose a time for the interview. Only five of the 10 contacted individuals responded to the message. And out of those five only three were interviewed. One of the individuals who responded at first had to cancel a couple times and then stopped responding. The other individual was the only lady in the whole group and never responded even after sending her a couple of reminders. None of the three interviewees currently work at the MoE.

**Procedure.** Results from the previous rounds of data collection informed the key informant interview protocol. After building the protocol (Appendix F), it was sent to the institutional review board at Western Michigan University to get their approval. Upon receiving the approval, the researcher started communicating with the key informants to schedule the interviews. Like the teacher interviews, these key informant interviews were done virtually, and audio recorded, upon the interviewees' consent. All the recordings were then transcribed using Sonix, a transcription software. All identifiers and personal information were removed from the
transcripts for confidentiality reasons and pseudonyms were given to each interviewee. The transcripts were then uploaded to the MAXQDA software program to be analyzed.

The Analysis of Qualitative Data

The analysis of qualitative data is a procedure that begins even before starting the data collection. For example, in interviews, the communication and accessibility of the interviewees could add to the interpretation of the findings of the study. Also, any incidents, discussed topics, issues related to the setting of the interview, the voice of the respondents, and the time and place where the interview was administered may also provide information of vital utility to the interpretations of the findings of the study. These, and other notes, are usually recorded by the researcher in what qualitative researchers refer to as memos (Creswell & Poth, 2018). Memoing in qualitative research provides the researcher with the ability of being reflexive and thoughtful during all phases of the study (Merriam & Tisdell, 2016). It is central to qualitative research to interact with all aspects surrounding the interview and the interviewee to be able to dig deep into the human experience (Birks & Francis, 2008).

For the current dissertation, the researcher applied thematic analysis as the main analytical technique to understand and make sense of the data. Braun and Clarke (2006) discuss what thematic analysis means. They refer to it as being “a method for identifying, analyzing, and reporting patterns (themes) within the data” (Braun & Clark, 2006, p. 79). In this dissertation, the researcher did not have a bounded theoretical framework among which the data was investigated. Instead, an inductive analytical approach was followed to gain more understandings of what the teachers and the key informants experience and believes are. Codes from the data were labeled, grouped, and reduced to more generic themes during an ongoing iterative process until the researcher reached a point where the data he collected was condensed in a more approachable
and understandable way, and there was not much information added. This is referred to as the saturation point in qualitative research (Nowell et al., 2017). These emerging themes were directly connected to the research questions in the results chapter.

Descriptive, in-vivo, and versus coding were partially applied to the interview transcripts. According to Saldaña (2016), descriptive coding is a basic type of coding that would help the researcher in organizing and summarizing the data. He describes in-vivo to honor the voice and expressions the respondents used during the interviews. Exact words and phrases are coded under in-vivo coding rather than generic meanings or interpretations.

As the current dissertation had two different sets of interviews (one for teachers and one for key informants), versus coding was applied. Saldaña (2016) describes versus coding as being “appropriate for policy studies, evaluation research, critical discourse analysis, and qualitative data sets that suggest strong conflicts or competing goals within, among, and between participants” (p. 115). The process of contrasting the thoughts and ideas collected from teachers with those collected from key informants helped in gaining deep understandings about the dynamics of the system and the points where the views of these two diverse groups of stakeholders misalign. The allocation of these points provided rich information for any development in teacher evaluation policies.

**Trustworthiness of the Qualitative Data**

As discussed previously, assuring the validity and trustworthiness of qualitative research is an ongoing process that the researcher should be able to engage in throughout the research process, legitimation as Onwuegbuzie and Leech (2006) calls it. Starting by the development of the survey, face and content validity were checked through sending the survey to teachers and evaluators to check if the items on the survey were related to teacher evaluation and the
associated constructs. For the interview data, the main ideas each teacher raised in his/her interview were resent to them for confirmation. This resulted in requests to reword some ideas in the transcripts. The main account for the trustworthiness of the interview results is the soundness of the results and its connectedness to the results of the other research strands. Overall, the findings from the interview did match the survey findings to a great extent. In some cases where there is a disparity between both, it will be mentioned.

**Limitations**

As Simon and Goes (2013) state, there is no human generated knowledge that has no limitations. It is thus the responsibility of the researcher to elaborate on what he thinks are limitations to his work. This is part of the reflexivity process and adds to the validity of the study (Creswell & Plano Clark, 2007).

For the current dissertation, there are several limitations which could be identified in relation to its design. First, as the researcher has earlier discussed, teacher evaluation is a complex topic. It would be hard to include all stakeholders in one dissertation. Thus, the focus was on getting the teachers’ and the key informants’ perspective. Other groups that could have been part of the study includes principals, supervisors, students, and parents. Second, the convenience sample which the researcher ended up having does not permit the generalization of the findings. The female teachers were underrepresented in the sample, in addition to teachers from some of the provinces. However, the gathered information is thought to be the first of its kind and that would help in understanding the teacher perspective on such an important topic.

Third, the contribution of key informants who currently work in the MoE in the interviews would have been of a great utility for this dissertation. It is not clear why those individuals decided not to participate, but it could be anticipated that their positions in the
ministry had influenced that decision. The aim was to get an insider view on what is currently being done in the ministry regarding teacher evaluation. Thus, some of the findings of the current study maybe irrelevant or already in the process of being applied. Fourth, the geographical context of the current study is Saudi Arabia, and any results may not fit to any other context. Lastly, it should be noted that this dissertation was carried out during the COVID-19 pandemic. The pandemic had an unprecedent effect on the function of academic departments. All advisory meetings, data collection, and interviews were done online. The education system and teacher evaluations were also done differently during this pandemic. The researcher had to remind all participants that the topic of the dissertation relates directly to teacher evaluations under the normal functioning educational system.
CHAPTER IV
FINDINGS

Teacher evaluation is thought to be one crucial step on the way to strengthen the quality of any educational system. The establishment and implementation of professional standards contributes immensely to the strengthening on evaluation systems. Using multiple data sources, this dissertation, investigates the alignment of the current teacher evaluation system in Saudi Arabia to the Saudi Teaching Standards (STS). Such alignment plays a key role in making clear to the education community what standards and criteria teachers are going to be evaluated upon. The dissertation also considers the distinct roles different stakeholders play in teacher evaluations. Qualitative and quantitative data were gathered to answer the research questions addressed in this dissertation.

In this chapter, the three research questions guiding this inquiry are answered. Table 3 is used as a guide to the layout of the chapter. The chapter is divided into five sections. The first section introduces the main characteristics of two of the best international practices in the field of teacher evaluation, which are the OECD teacher evaluation framework, the Danielson group teacher evaluation framework and the InTASC. The second section provides a description of the teacher evaluation system in Saudi Arabia. This description relies on the teacher survey data as well as the interviews with teachers and the original publication of the STS. In the third section, a comparison between the Saudi teacher evaluation system and teacher evaluation systems from some of the OECD countries is made. The fourth section provides the results of the alignment between the evaluation system and the STS. Finally, in the fifth section, the misalignment between both is highlighted, with steps needed to move forward. For each of these sections, different lenses are utilized to explore the teacher evaluation system in Saudi Arabia.
key informants, and international best models and practices collectively aid in reaching an objective assessment on what is happening and what needs to be developed.

**Document Analysis of the Best Practices of Teaching Standards**

The Saudi Teacher Standards (STS) (Appendix B) that was published in 2017 will be compared to the Interstate Teacher Assessment and Support Consortium (InTASC) standards and the Danielson Framework for Teaching (FFT). These frameworks were introduced in the second chapter of this dissertation in (see Tables 1 and 2). The InTASC model provides general dimensions that are required in the teaching profession, with several descriptive performances under each domain that teachers should demonstrate. The STS, on the other hand, is a more detailed document, domains, standards, and sub-standards. Nonetheless, comparing the two shows that all the domains in the InTASC are covered in the STS.

The FFT is a more detailed document in its domains and subdomains than the InTASC. Contrasting the FFT to the STS reveals that both documents share almost the same standards. However, the STS is more detailed on the professional knowledge of a teacher. Also, the FFT tends to be more procedural in the language used to describe its instructional method. For example, under the instruction domain, the FFT sets five procedural standards such as the use of questioning and discussion techniques. In contrast, the STS provides a general statement using variety of instructional methods.

Given this short contrast between these documents, it is obvious that the authors of the STS reviewed the relevant literature in producing the standards. The similarities between the documents are large, and the few differences could be related to the different contexts and/or the differences in the overall goal of the education system. It is important to note here that it is not
the aim of the current study to critically discuss the STS, rather the aim is to contrast it with similar documents that have been around in the literature for a relatively long time.

**Teachers’ Perception of the Saudi Teacher Evaluation System**

The teachers’ survey data, in addition to the 11 interviews with teachers, provided a detailed view on aspects of teacher evaluation in Saudi Arabia. In this section, the results from these sources are utilized to describe the current practice of teacher evaluation in Saudi Arabia. Significant differences between sub-groups from the sample on four indices (Working Conditions, Standards-Practice Alignment, Administrative Change, and Skills Change) are also reported. This will help understanding how diverse groups of teachers view the evaluation system. It is important to note, however, that there were a small number of significant differences and that the study did not intend to compare groups. Instead, the researcher thought the comparison might contribute to understanding teacher evaluation in the country. Before getting to the aspects of teacher evaluation, a brief description of the sample of teachers who completed the survey is provided.

The original number of teachers who started the survey was 704 but there were several who did not agree to consent to the process. Others left most of the questions unanswered. Hence, the total number of teachers who completed the survey was 642. Amongst those 414 (64.5%) are males, and 228 (35.5%) are females. All 13 provinces were represented in the sample but with various levels of representation. The largest represented province was the province of Makkah with about 31% of the sample. Riyadh province comes second with about 24.6% of the sample. The eastern province follows with 18.8%. The least represented province was Najran, with about 0.8%. The average years of experience was 16.6, with 20 years of experience being the mode of the sample. This indicates that the sample of teachers are on the
older side, as the average age for those participating which was 41.6 years. The average number of schools where the teachers worked was about five schools, which also aligned with the samples’ high age and experience. In terms of the subjects taught by the teachers, the most common were science, religion, Arabic language, and math, respectively. Most of the sample (42.8%) were high school teachers.

Four indices were calculated from the teachers’ responses to the four points Likert type items (1 = Totally Disagree to 4 = Totally Agree) on the various parts of the survey. The indices are: Working Conditions, Standards Alignment, Administrative Change, and Skills Improvement. Appendix G includes all the tables for the comparison analyses that were conducted.

**Table 8**

*Descriptive Statistics*

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>22</td>
<td>60</td>
<td>41.62</td>
<td>6.56</td>
</tr>
<tr>
<td>Years of experience</td>
<td>1</td>
<td>34</td>
<td>16.60</td>
<td>7.07</td>
</tr>
<tr>
<td>Number of schools</td>
<td>1</td>
<td>20</td>
<td>4.99</td>
<td>2.84</td>
</tr>
</tbody>
</table>

The teachers’ interviews, as described in the previous chapter, were done to delve deep into some of the survey findings and search for explanations to them. There were six female teachers, and five male teachers who agreed to take part in these interviews. Eight different provinces were represented in this small sample of interviewees. After going through the interview transcripts, themes started to emerge from the data files. The first round of coding resulted in 21 themes that had a total of 428 codes in them. After a while, these themes and codes were revisited for a second round of coding. After this second round, the researcher made some
changes and merged some of the themes into others, resulting in eight main themes and 13 subthemes. Figure 4 shows the way these themes and subthemes connect to each other. During the presentation of the results, the quantitative results are presented first, then the related theme from the qualitative analysis are presented, with relative quotes from the interviews as appropriate.

**Figure 4**

*Themes and Subthemes Layout*

---

**Teachers’ Working Conditions**

The average number of weekly lessons teachers must teach was 18 lessons. The mean of the Working Conditions index was 2.18. There were no significant differences between males and females on that index. However, according to the post-hoc ANOVA analysis (Appendix G), there is a significant difference between math and science teachers and between science and Arabic language teachers. High school teachers also differed significantly from middle and elementary schools’ teachers on the Working Conditions index. The average number of hours
teachers spent preparing for lessons was 7.6 hours per week. There were about seven out of the 11 interviewees who mentioned that the overload work, especially in non-instructional activities, is overwhelming. Sara for example reported in the interview with her, “there is a lot of stress, and it does affect the way we are evaluated.” Khalid tried to explain the nature of the stress he faces by relating it to the lack of support from other school personal and he also asserts that the amount of work teachers do is not reflected in the current teacher evaluation system.

Figure 5 summarizes the results of the question that was concerned with some aspects of teachers’ working conditions. Class size is an issue when it comes to the efficiency of time as about 70% of the sample thinks class sizes does not allow teachers to have the time to meet all students’ needs. Protection of teachers from extra work that is not related to the essential role of teaching was also an issue that is related to teachers working conditions. About 76% of the teachers think they are not protected from extra unrelated work. The most agreeable item in relation to teacher working conditions was the non-instructional time provided for teachers to prepare for lessons and conduct non-instructional work. Teachers in the interviews and in the survey of open-ended questions raised the issue of large class sizes frequently. Assigning duties that are not related to instruction was another main idea that was raised under the theme of challenges. Rahaf talks about the overwhelming number of tasks a teacher is responsible for by saying, “More than half my time during a school day goes into planning and executing some non-instructional activities, and when the time comes to execute my main job, which is teaching, I do not have any more energy to do so.” She was prompted by the researcher with a question about who asks her to do these extracurricular activities and if she has the right to deny them. She replied:

It is the principal, and she is so invested in the extracurricular activities as those are the ones that would be reported to the Ministry, and it will show how she is
doing an excellent job. So, it is a priority for her. The student learning comes later her priority list. And of course, I cannot reject accepting these tasks as she is the one who will evaluate me, so I better accept whatever she asks, to some extent.

**Figure 5**

*Teaching Working Conditions as Reported by Teachers*

Teacher Evaluation

The core of this dissertation is about teacher evaluation. In this section, the practice of teacher evaluation will be described according to the teachers who completed the survey and participated in the interviews. Regarding the frequency of the formal evaluation teachers receives, most teachers choose *twice a year*, which was explained in more detail in the interviews. Under the "inconsistency" theme, five of the teachers described the inconsistency in the number of times they get evaluated. Fahad, for example, reports that the frequency of
evaluation depends on the superintendent. In some locations, the superintendents are overwhelmed by so many teachers that they concentrate more on the new teachers in their administration zone. Older teachers who have been evaluated by the same superintendent in the past semester or academic years may not have an evaluation visit if time does not allow for it, and the previous evaluation will be kept in the record without any updates.

Informal evaluations are not part of the system according to about two-thirds of the survey sample. In the interviews, the idea of informal feedback and support from other teachers was discussed. The responses varied to an extent. However, it seems that the MoE had a specific initiative to motivate teachers to visit each other and to indulge in discussions about their practice and how to improve it, but it was not applied as intended. Rana, for example, says that these informal peer visits are just done to fill in some paperwork and no teacher benefits from it. She explains it would be weird to correct a teacher or criticize the way they teach. On the other hand, Sara thinks these visits do happen sometimes, and when they happen, they are beneficial.

**Figure 6**

*How Many Times Do Formal Evaluations Occur?*

![Pie chart showing the frequency of formal evaluations](image)

- Red: Once every two years (3.9%)
- Blue: Once a year (36.2%)
- Green: Twice a year (50.9%)
The teachers were asked to specify the contribution each stakeholder group has in their professional evaluation. It could be seen from Figure 6 that principals and superintendents are solely responsible for the evaluation of teachers. Peer evaluation is minimally practiced according to 35% of the teachers. Under the theme “evaluators,” many teachers expressed how principals are in total control of their evaluations, despite having the superintendent as an evaluator as well. Essentially, the final decision is in the principal's hands. This is upsetting for most teachers, as described by Khalid. He talked for a while about how inexperienced the principal could be in evaluating teachers, especially those who are not from the same field. For example, a principal who used to be an Arabic language teacher and now is a principal would be responsible for evaluating teachers of other subjects like math and science without any prior knowledge about these subjects or how they should be taught. Sara and Rahaf both talked about how subjective the evaluation becomes when the principal does the whole evaluation. They both mentioned stories about them being in schools where principals were giving them positive evaluations and as soon as they moved to other schools the entire system of evaluation changed and they were then evaluated lower than they used to be due to them not being friends with the principals.

Another idea that is related to the subjectivity of the evaluators and was dominating the "challenges" theme is the idea of the subjective grading criteria. All teachers who were interviewed have stated that they have never encountered a teacher who had an evaluation score below 90 points out of 100 points. And the average score is above 95 points. These scores mean nothing, says Saad, “it is just arbitrary numbers.” In fact, Fahad talks about how principals might use these scores to support teachers they do not like in their desire to relocate, as the higher their evaluation score is the faster their relocation request will be fulfilled.
All the teachers in the interviews pointed out some notable examples of principals who were so transparent and honest in conducting the evaluations; however, they also point out that these are not the norm. Hamad described how the system is not helping principals to be good evaluators. He spoke:

Principals and superintendents are not to be blamed solely when it comes to the broken evaluations we get. The entire process and policy should be changed. What are the consequences of these evaluations? How principals and teachers are going to take evaluations seriously if they already know that it has no consequences.

Figure 7

*Stakeholder Involvement in Teacher Evaluation*

Regarding the source of data used in the evaluation of teachers, the teachers were asked to specify the extent different data sources were used in their professional evaluation (see Figure
7) Teacher attendance and lesson plans seem to be at the top when it comes to gathering data about teachers. Classroom observations, participation in extracurricular activities, and student achievement seem to be widely considered as well. In the interviews, all teachers pointed to one exclusive instrument that is used to collect data on teachers' performance, which is the evaluation form (Appendix A). According to the teachers, this form is meant to be filled in by the principal and the superintendent during a classroom visit. There is an administrative part to the form which considers attendance and cooperation with the school administration. The principal of the school has about 80 percent of the overall score and the remaining 20 percent is in the hands of the superintendent. However, eight of the teachers reported that the last say in the overall score is for the principal.

The teachers did have mixed opinions on how just and valid the evaluation form is. An extreme example is Ali. He was so direct in referring to the form as being unjust and believed that the ministry should stop forcing its use. Arwa finds the form acceptable, but she thinks the way it is applied is so subjective; hence, the opposition to it. She, and others, bring up the concept of nepotism, where the circle around the principal always gets the best evaluation scores even without being observed in the classroom. To that, Hamad had an incident where he was new at his school and the principal of the school always gave relatively low scores in the evaluation to new teachers in his school. Fahad says:

I knew that he gave me a low score just because I am new in high school despite being better than most of the old teachers in that school. My load was even higher than most of those who were given higher scores. So, I went to him and asked why my score was less than others. He was not expecting me to argue with him about it and put it solely on the superintendent without providing any feedback or evidence. It was not that a higher score will provide me with anything special, but I just felt being treated unfairly.
A less extreme opinion on the evaluation form comes from Rahaf, a science teacher who taught in about five schools across the country. She thinks that the form is just and comprehensive of what a teacher should be doing but she still sees a huge disparity in the application of the form. She states “the problem is from the principals who are not trained well enough to conduct the evaluations. Nepotism and corruption are what is making the whole evaluation system useless, not the form itself.”

**Figure 8**

*Methods of Evaluation Used in Evaluating Teachers*

The feedback a teacher receives after the evaluation is crucial in the process. The teachers were asked whether they received such feedback or not. About 70% of the teachers reported that they do get feedback after the evaluation. However, about 35% of those who did receive
feedback reported that it is exceedingly rare that they get a chance to discuss the feedback they receive with the evaluator (Figure 8). Ali, in his interview, reported that there is nothing in the policy of teacher evaluation that requires the principal or the superintendent to sit with the teacher and discuss his or her evaluation report. Hind links it to the subjectivity of the principal. If he or she is open to having a discussion with the teacher they would plan for such meetings, but no one cares if it happens or not. Khalid also mentions that most of the time the evaluation report and score are not released by the principal until the end of the academic year, which makes any feedback or discussion of no utility.

**Figure 9**

*Evaluation Feedback*

---

**Do you get feedback?**

- Yes: 69.2%
- No: 30.8%
The survey asked the participants to rate their awareness level of the teaching professional standards. About half the teachers were not aware of the details of the standards at all (see Figure 11). For those who were aware of the standards, the researcher asked them about how they knew about the standards and most of them knew about it by self-research or through colleagues (see Figure 12). These findings align perfectly with the responses of teachers in the interviews. While talking to teachers about the standards it was obvious that many of them were not fully aware of them. The researcher had to explain what he meant by the standards in diverse ways before teachers knew what he was referring to. Under the “standards impressions” subtheme, five of the teachers reported clearly that they have never seen the standards. Four other teachers reported that they know the standards exist but never looked at them. Just two of the teachers reported that they have read the standards. Rana did not know about the standards until the ETEC (Education and Training Evaluation Commission) announced that the teacher
licensure program and examination will be based on the standards. Fahad and Saad both think that teachers are overwhelmed by many regulations and policies every day and that they are not willing to engage or participate in new policies or procedures they never participated in creating. Rahaf explains that the miscommunication between the teachers and the ministry is not just about the standards. She states that teachers are usually the last group in society to know about some major changes in the way they should work and teach.

**Figure 11**

*Level of Awareness of the Teaching Standards*
The core of this dissertation was covered in a Likert-type question about the level of coverage each of the main STS received in the evaluation of teachers. Figure 13 shows that most of these standards are reasonably covered in the evaluation of teachers. The average overall coverage of these standards was 3.36, which is high. An independent samples $t$-test revealed that female teachers differ significantly from male teachers on their view about the representativeness of the standards in the current evaluation system (Appendix G). A simple linear regression showed that age is a significant predictor of the index standard coverage; however, it does not explain much of the variance ($R^2 = 0.01$). The least covered of these standards is the amount of professional training the teacher had taken and the promotion of cultural diversity; however,
looking a little deeper into the thematic analysis of the teachers' interviews, it could be argued that these survey quantitative results need more explanation. It needs to be said that as most of the teachers were not familiar with the standards and with the limited time available for the interview, the depth of the discussion about these standards was not satisfying. However, some of the discussions that took place may be useful.

Under the “standards” theme, teachers generally reported on the alignment between the current evaluation system (i.e., the evaluation form) and the STS in general. There were some standards where some of the teachers think it is not fully covered and these are the promotion of cultural diversity, the amount of professional development the teacher had, and the relations with other teachers and with the community. Several teachers also think that much of the success the teacher achieve through participating in extracurricular activities is not reflected in these standards, nor in the current evaluation system. Ali and Hind both think that more recognition is highly desired, especially from teachers who work so hard in lacking environments.
The consequences of teacher evaluation were also addressed by asking teachers about the influence their professional evaluation had on them. From Figure 14, it is obvious that most teachers think their evaluation results did not have any influence on their employment status, financial outcomes, or responsibilities. It was emphasized by many of the teachers in the interviews that teacher evaluations are done as a routine task and there are no consequences for them; hence, the unseriousness about them. Ali, for example, responded to the question of whether there are any consequences for the evaluation, in a noticeably short simple answer by saying “nothing at all.” Fahad mentioned that if a teacher’s score was less than 70, he or she would not get the annual raise at the end of the year, but he stresses that it never happens. Rana
mentioned that sometimes the superintendent suggests resources if the teacher is lacking in some areas, but this does not happen all the time.

**Figure 14**

*Personal Consequences of Evaluation*

The teachers were asked about the impact evaluation had on their instruction and classroom management skills. Figure 15 shows that for most of the listed skills, about half the sample thinks their evaluations did have an impact. The least impacted skill is the teaching of students with special learning needs. However, when teachers were asked about this in the interviews their responses varied. Some of them think that it is upon the teacher, whether he or she is ready to develop and act upon the evaluation findings or not. As previously noted, teachers in the interviews questioned the credibility of the whole evaluation system being beneficial in
some ways. Nora emphasized that no one should assume that a teacher who does not believe in a
certain way of evaluation would have respect for its output, if there were any at all.

**Figure 15**

*Consequences for Evaluation of Teaching Skills*

Teachers were also asked about their perception of the overall evaluation process and
whether they felt it was for summative or formative purposes. More than 78% of the sample
thought that the evaluations they went through are mainly for summative purposes (see Figure
16). Under the "methods" theme, four teachers were inclined to judge the evaluation system as
being summative, whereas the rest of them were even hesitant to choose between formative and
summative as they think the evaluation system is useless and represents neither of them. Fahad,
for example, says that "what does it make of a difference if I say it is formative or summative, it

<table>
<thead>
<tr>
<th>Concerning your formal evaluation, to what extent did it directly lead to the following:</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.1 Your classroom management practices</td>
<td>38.0</td>
</tr>
<tr>
<td>24.2 Your knowledge in your main subject field</td>
<td>36.0</td>
</tr>
<tr>
<td>24.3 A development or training plan</td>
<td>36.0</td>
</tr>
<tr>
<td>24.4 Your teaching of students with special learning needs</td>
<td>36.0</td>
</tr>
<tr>
<td>24.5 Your handling of students disciplines and behavior</td>
<td>35.0</td>
</tr>
<tr>
<td>24.6 Your teaching of students from different backgrounds</td>
<td>36.0</td>
</tr>
<tr>
<td>24.7 The emphasis you place on improving your students test scores</td>
<td>30.0</td>
</tr>
<tr>
<td>24.8 Your students assessment strategies</td>
<td>30.0</td>
</tr>
<tr>
<td>24.9 Use of different teaching methods</td>
<td>30.4</td>
</tr>
<tr>
<td>24.10 Handling of students problems and needs</td>
<td>31.0</td>
</tr>
</tbody>
</table>
is useless after all!” Hind describes that by saying, “the main intent of the system is to be summative as it attaches a score to each teacher, however, this score is meaningless and useless at the same time.”

**Figure 16**

*Evaluation Type*

![Evaluation Type Diagram]

**Teacher Professional Development**

Teacher professional development should be connected to the evaluation results of teachers for it to be most efficient and useful. Teachers were asked about the number of professional development programs they are required to attend per year as part of their evaluation process. Responses to this question varied to a great extent. About 35% of the teachers think they are not obligated to take any training programs. Another 35% reported that they must take between 3 and 5 professional development courses per year and that this
professional development is considered as part of the formal evaluation they receive. The remaining 30% of the sample thinks teachers are required to attend more than three professional development courses each year (see Figure 17). More than 70% of the teachers reported that they do attend professional development that is not imposed by the schools or the Ministry. More than half of those teachers reported that they pay for these development courses out of pocket.

In the interviews with teachers, they elaborated more on professional development. Most of the teachers described how important professional development is in keeping them updated on innovative instruction and learning. It was clear from the interviews that teachers are not obligated to attend any professional development. Saad had the evaluation form and pointed out what they are on and that there was no mention of training or professional development in the form. Fahad talked about some improvement that is happening currently where professional development is going to be part of the evaluation system, but it has not yet been applied. Sara and Khalid both mentioned attending many training courses on different topics but they both feel that the value of most of the courses they attended is minimal. Five teachers raised a crucial issue about how principals select them for training courses. Hamad describes his experience by stating:

The process starts with the local administration office asking school principals to select teachers from specific training courses. Principals do not have any official criteria to choose upon. And remember that training takes place during the school day so the principal will need to find someone to cover for any teacher who is going for the training and most of the time they will just select someone who does not teach the main subjects as those will leave behind fewer lessons to be covered. In other cases, these programs are used as an incentive by principals for teachers who cooperate with them and do what they are told. Very few principals, to my knowledge, will select teachers in real need for these opportunities to develop.

Teachers also expressed the absence of alignment between what professional development they need and what the ministry offers. Ali states, “we, as teachers, do not have a say on what we actually need from professional development, and in most cases, it is repetitive,
and we look at it as a relaxing out-of-school time.” Rana had an interesting story where she was a high school teacher for 7 years and was then sent on a scholarship training program to the United States. When she went back to her school, she was directed to teach at the elementary level, as there was a shortage of teachers there. The interesting part is that this was not considered in her evaluation and she was not provided with any support on teaching in elementary school.

It should be noted that some of the teachers have mentioned that they have benefited from some of the training and professional opportunities. Hind, for example, a religion teacher, has been to a 3-day training on engaging students in religious classes. Hind reports that the training was more of a community of religious teachers who shared their experiences and discussed them, and that she gained a lot from those who are more experienced than her in this field.
Another aspect of professional development that gives an indication of how much it is related to the evaluation process of teachers is the way the teachers get to choose the professional development courses they attend. About 75% of the teachers reported that they choose the professional development courses they want. The teachers in the interview described how principals select teachers to attend professional development based on their subjective decisions and how most of the time if you are in a good relationship with the principal, you will end up being selected to attend many of these professional development activities. It should be noted how the process is seemingly getting developed. Hind described a new process where the local administrative office lists all the training they offer on a website and the teachers will log into their accounts and choose whatever they think fits their needs.
The types of professional development courses did vary according to teachers (see Figure 18). In-person and virtual training courses are the main type of professional development. Peer visits come third. In the interviews, teachers mentioned that some principals encourage the practice of peer visits, but the procedure was not relevant. Rana describes the practice as calling out co-workers for their practice in the classroom instead of it being a positive interactive process. She says, “I would never want to fill in a form about my co-worker practice with any negative impressions about their work.” Saad elaborated more on that by saying, “Usually teachers will fill these forms in the teachers’ office collectively with all positive comments, so no one gets any bad impressions from the administration.”

Figure 18
Types of Professional Development
Teachers as a Professional Community

The survey asked several questions about teacher engagement in school administration activities and creating policies. The aim was to assess how involved teachers are in the school community and if they have a voice as a professional community. Attending staff meetings was amongst the most endorsed activities. Professional visits to other teachers were also highly endorsed. The teachers in the interviews talked a little about the support they provide to other teachers in their schools and get from them. Sara, however, mentions that any kind of such interaction between teachers is done individually most of the time and not in an organized way. Fahad and Ali both advocate the use of social media and technology to create interaction between teachers of the same subject across the country. This would, as they describe, enhance the teaching of all the involved teachers.

Figure 19

Teacher Engagement in Professional Community Activities
Comparative Analysis of the Saudi Current Teacher Evaluation and the OECD Best Practices

In this section, a comparison is made between the practice of teacher evaluation in Saudi Arabia and across some of the OECD countries. The comparison is based on the following main themes: the existence and framework of teacher evaluation, the procedures of teacher evaluation, handlers of teacher evaluation, and the consequences of teacher evaluation. The indicators in Table 6 will also be covered in this narrative comparison.

Existence and Framework of Teacher Evaluation

The existence of teacher evaluation is a trait of high-performing countries on the 2015 Program for International Student Assessment (PISA, Avvisati, 2018). In any way it occurs, teacher evaluation ensures that teachers can effectively teach deliver in the classroom. Nusche (2013), states that 23 out of 29 OECD countries have a rigorous teacher evaluation framework that covers all theoretical and procedural aspects of the system. Those countries differ in the way teacher evaluation is executed, but that does not affect how well-performing the teachers in these countries are (Avvisati, 2018). According to the data collected for this dissertation, teacher evaluation is practiced in Saudi Arabia, and teachers do encounter performance evaluations throughout their careers as teachers. However, the manner in which teacher evaluation is practiced is not connected to a comprehensive framework that details the theoretical basis of the system and how it should be executed. According to Ali, who used to be a school principal for more than 10 years, teacher evaluation is based on the circular letters the Ministry sends occasionally to schools. He states that he used to have a folder full of these letters and had to go back and look manually at them to know how he should evaluate teachers. In some cases, new letters contradict old ones, causing the Ministry to send an explanation of how to deal with contradictions.
In terms of teaching standards, the OECD’s suggested practice for teacher evaluation stresses the importance of establishing teaching standards for an improved and effective teacher evaluation system (Avvisati, 2018). It is also recommended that these standards are developed with the participation of as many teachers as possible, through union representation or mass surveying (Nusche, 2013). Nusche (2013) argues that teachers should own these standards and believe in them. If so, teachers will be more cooperative and open to being evaluated, which is a crucial element in the effectiveness of the evaluations. The ETEC has indeed accomplished this step by publishing the STS in 2017. Both the survey data and the interviews with teachers show that the standards are currently a fact. However, an overwhelming majority of teachers have not looked at the standards or even heard of them, much less been a part of their development. A narrative of opposition to the standards was clearly manifested in the “standards” theme from the interviews with teachers. For example, Ali states, “I have not read the standards and will not read them, why should I when we have not participated in their development.” Arwa states that about half the teachers she knows do not support the development of such standards, as they will increase the burden that is put on the teachers’ shoulders. It was also observed that the MoE and the ETEC have not put enough effort into raising the awareness of teachers about professional standards. The survey data shows that most of the teachers who are familiar with these standards have developed this awareness through their own research and reading. As in most of the widely used teaching standards, the STS cover the areas of professional responsibilities and ethics, professional knowledge, and professional practice.

**Procedures of Teacher Evaluation**

According to the OECD (2015) Education at a Glance book, teacher evaluation happens periodically in most of the OECD countries, especially those with high performance on the
PISA. A compulsory process of formal evaluation is usually done on an annual basis (Isoré, 2009). Different countries have diverse types of evaluations such as regular evaluations, evaluations for completion of probation, and evaluations for registration. These can appear collectively in some education systems but in most countries, at least two of these are in place (OECD, 2015). For the Saudi education system, teacher evaluation does occur biannually, once every academic semester, and the only type of evaluation is the regular evaluation that all teachers should receive. Probation periods do not exist in the Saudi education system; however, Khalid and Ali in the interviews have mentioned that principals and superintendents are usually more attentive to evaluate new teachers. The more the teachers are experienced, the less they will be evaluated.

Multiple methods of gathering data related to teacher performance is a practice in most high-performing OECD countries (Avvisati, 2018). Nusche (2013) asserts that for an accurate and fair teacher evaluation, data should be gathered from multiple sources. In the context of Saudi Arabia, the survey results show that multiple sources are utilized for data gathering. These sources include classroom observation, lesson plans, and student scores. Student scores are mentioned when talking about teacher evaluation; however, Nusche (2013) points out that this method is not the norm in most of the OECD countries despite the push in many countries to adopt it as a method. In the interviews in the current study, the teachers indicated that this source is not widely used. Fahad, for example, talked about using student achievement in the evaluation of teachers by stating, “we do not have standardized tests in the schools and the principal does not have access to any of the students' scores on the tests that teachers prepare.” Ali describes how student achievement is part of the evaluation by stating:
During the classroom visit, the principal will ask a couple of students some questions about the content being taught in the class. If they answer correctly, the teacher will be given a high score on the student achievement indicator and if they do not answer then the teacher will get a low score.

The teachers also noted the subjectivity of the evaluations they received and how these evaluations were determined solely by the principal without even conducting a classroom visit in some cases. In such circumstances, it is hard to see use of multiple sources of data, as it all goes back to the principal subjective judgment. For that, Hamad states, “I have not seen a principal or a superintendent in my classroom for the past 6 years. I always get the evaluation report at the end of the year, and it is always a high score.”

Despite all the efforts to increase the number of data sources regarding teachers’ effectiveness, classroom observations remain the leading source in about all the OECD countries (Isoré, 2009; Nusche, 2013). Most of the high-performing countries on the PISA do combine teachers interviews with classroom observations, especially at the end of the probation year (OECD, 2015). This leads to the importance of the feedback teachers get after being evaluated. Systems that support the formative aspect of teacher evaluation will always have a mechanism for providing feedback to teachers after their evaluations (Nusche, 2013).

In the Saudi educational system, classroom observations are the main and only source of data that is collected to evaluate teachers. Feedback turns out to be an activity that is required to be done by principals after each evaluation, as indicated by the survey results. A third of the survey sample almost never had a meeting to discuss the feedback with the evaluator. The interviews shed more light on the process of getting feedback and being able to discuss it. Nora, for example, states that the usual practice is for the principal to hand the completed evaluation report to the teacher and tell them to stop by their office if they have any issues with it, but this does not happen most of the time. Fahad reports that the only reason a teacher would confront or
discuss an evaluation report with the principal is when they have received a relatively low score. He also describes how he has never seen a teacher getting a score less than 90 in his whole 15 years of being a teacher. These are procedures that have no meaning for most of the teachers, he admits.

**Evaluators**

Stakeholders of any education system are several and each has their own interest in ensuring that the system is effectively working and that teachers are doing what they are supposed to do. The most common involved entity of stakeholders in the evaluation process are school principals (OECD, 2015). Superintendents, school inspectors, and peers also carry on some evaluation responsibilities in some countries. Avvisati (2018) stresses on the importance of having more than one stakeholder group involved in evaluating teachers, as this will significantly increase the level of accuracy and integrity of the system. He recommends having regular monitoring and developmental evaluations throughout the year that are done by competent internal evaluators. These tools are useful in tackling any shortage of skills or knowledge that might affect any of the teachers’ performance. The second stakeholder group he recommends is the government or the authorities represented in the inspectors or superintendent. For those groups, the evaluation will usually focus on the summative part of the evaluation system and intends to establish a basis for career progression regulations. Before any of those stakeholders are involved, it is crucial to ensure any evaluator is competent and capable of executing the evaluation in a professional and honest manner. Untrained and incapable evaluators who are not aware of the importance of the task they are taking usually cause the system to be prejudiced and dishonest (Nusche, 2013).
In the Saudi system, and based on the data gathered for this dissertation, school principals and superintendents are the only two evaluators, and the principal holds about 80 percent of the total score. That is tremendous power and authority given to principals, according to eight of the interviewees. Hamad states that the amount of power the principal has in the evaluation of teachers forces all teachers in the school to be cautious and try to build good relationships with them, as, “If you are close to the principal, you will be treated greatly in all aspects of schooling but if he does not like you then you better find another school.” Saad also talks about having to move to a different school just because he did not have a good relationship with the principal in the previous school. Superintendents are also part of the evaluation system, but they are usually less powerful, as the final word is usually given by the principal of the school; however, about five of the interviewees think superintendents are better at conducting the evaluations, as they usually give feedback that is helpful to the teacher. Rana indicates that the visit of the superintendent is usually useful, and they are better, in general, than the school principals.

To find out why principals were usually unfavored as evaluators, a question was asked to the interviewees about the traits they dislike in an evaluator. All interviewees mentioned the lack of experience in conducting evaluations. Ali, who used to be a school principal, admits that while he was a principal, he used to evaluate teachers unprofessionally. He states, “I became a principal just because I was the only teacher in the school who had been in the same school for more than 5 years. I did not have any experience or preparation of becoming a principal and in evaluating teachers particularly.” Khalid mentions that recently the Ministry began to offer training for leaders, and that in these training courses there is some content about how to evaluate teachers; however, these are not mandatory, as he states.
The teachers were also asked about their thoughts in including more stakeholders in the evaluation process such as students and parents. The responses varied. Some think that the students and parents are not going to provide honest opinions about teachers they do not like and that it would be extremely hard to rely on them in that matter. Others think that parents and students might provide secondary information about the teacher’s performance that might be useful for development but not for deciding on a teacher’s proficiency level.

**Consequences of Teacher Evaluation**

Resolving the tension between summative and formative aims of an evaluation system is crucial for the system to be efficient and useful (Avvisati, 2018). Different stakeholders aim for different outcomes of the evaluation system. The authority, or the government, is more interested in the accountability aspect of the evaluation, whereas the teachers are more about the developmental aspect of the system. Most of the OECD countries account for both aims, but the fact of the matter is that one of these aims usually gets more attention from authorities. This would not be so problematic if the evaluation system was feeding the developmental aspect of the system. As Avvisati (2018) indicates, a characteristic of the leading education system amongst the OECD countries on international assessments is an evaluation system that feeds into the development of teachers and schools. The more the evaluation is isolated from the developmental aim, the less the utility of that system. This link starts at the planning phase for professional development activities and training courses. This starting point should always be included in the evaluation results to allow for any gaps or weaknesses that need to be addressed. Another observation in 11 OECD countries is that they link the evaluation system to a reward and incentive scheme. For the summative aim of the evaluation, regulations on career advancement regulations should be clear and well designed.
In the context of Saudi Arabia, according to the data gathered for this dissertation, it could be recognized that the evaluation system is useless in general. First, there is no official link between the evaluation outcomes and professional development. From the survey results, one can see that about 80% of the sample think that the evaluation results did not affect the professional development they receive or engage in. The same trend was observed during the interviews with teachers. Hamad talked about the procedure a teacher follows to get professional training. He states, “It has nothing to do with the evaluation of teachers. It solely depends on the available training courses. You could be skillful in all what is offered but still go and attend something that will not benefit you.”

Saad also echoed what Hamad stated and added that the professional training that is offered has not changed in a long time and it is highly likely that you will be sent to attend a training that you already attended in previous years. Rahaf, on the other hand, talked about the superintendent recommending professional development opportunities after she observed her in the classroom as part of her evaluation. She states, “It was a wonderful opportunity and I think the superintendent was very helpful in that matter but unfortunately that does not frequently happen.”

Similar findings occurred for the summative aspect of the evaluation. The survey results show that most of the teachers think their evaluations did not have any effect on their career development, nor were they given any recognition for getting good scores on their evaluations. The interviews have also pushed the same narrative. Fahad puts it clearly that the current evaluation system is a waste of time and money. Hind stresses that the lack of clarity and honesty in the evaluation system has led teachers and principals to just play the system and do what they
are ordered to do in the minimal effort they can just to get rid of the burden because they all know that nothing will depend on that evaluation.

The Alignment Between Teacher Evaluation and the Saudi Teaching Standards (STS)

In this section, the aim is to match the STS to the current evaluation procedures of teacher evaluation and to indicate what teaching standards are not addressed. The current practice will be represented by the items on the formal teacher evaluation form (Appendix A), whereas the STS are shown in (Appendix B). The sections below start by introducing the overall match and mismatch between both documents. Then, each standard from the STS is visited and data from the survey and/or the interviews is provided to show if that standard is represented in the current practice or not.

The current evaluation form consists of 19 items, whereas the teaching national standards have 10 primary standards and a total of 32 secondary standards. There are 17 secondary standards that are represented in the evaluation form, whereas 15 are not. In general, most of the items on the evaluation form are generic and subjective. There is even a double-barrel item that asks about two different concepts at the same time (item 7). The STS are more specific in wording and tend to state practices rather than general concepts. It is worthwhile to note that the national standards document has an extensive explanation for all the indicators of these standards; however, the focus of this dissertation is on the alignment between the STS and the current teacher evaluation practice. Hence, these extensive indicators will not be included.

The first primary standard in the STS is adherence to Islamic values and professional responsibilities and strengthening the national identity. Two of the secondary standards under this primary standard are not covered in the evaluation form, adherence to Islamic ethics and the strengthening of nationalism and cultural diversity. The third standard, committing to
professional ethics, is addressed in item 14 on the evaluation form. From the survey data, it can be seen in Figure 13 that most of the teachers think that adherence to Islamic values and the promotion of national identity are both covered in the evaluations they went through. Khalid, in his interview, explains some of this when he talks about issues that go without needing to be evaluated. He talked about the Islamic values and the concept of nationalism by stating, “it does not make any sense to evaluate a teacher on how they adhere to the Islamic values as we all are practicing Muslims. Same goes to the love of our nation.” A smaller but still significant proportion of teachers also think that the evaluation process they encountered covers the concept of promoting cultural diversity. However, as pointed earlier, it is not clear how this concept is covered in the current evaluations, as the evaluation form does not have any items addressing cultural diversity. Adherence to the professional responsibilities was highly agreed upon amongst respondents, as about 88% of the sample think this standard was important in the evaluation they encountered. Eight of the teachers in the interviews mentioned that professional responsibility is amongst the most prioritized concepts in teacher evaluation—that the teachers adhere to what they are told to do. It is also highly reflected in how the interviewees describe the significance of their relationship with the principal and do what they are told in their evaluation.

The second primary standard relates to professional improvement where there are two secondary standards; that is, having a plan for professional development and improving professional performance. Both standards are not reflected in the current evaluation form. In the survey, teachers responded to a question of how many professional development courses they should attend each year, and the responses varied. Yet, more than a third of the sample think they are not obligated to attend any professional development. The interviewees described how professional development is not obligatory but recommended. Ali, for example, states that
teachers are usually nominated by the principal to attend training courses and in some years some teachers will not be nominated either because there is no planned training or merely because the principal does not want to nominate them. Regarding the second secondary standard related to improving professional performance in accordance with the professional standards, it is obviously not reflected as these standards are new and not linked to the teacher evaluation practice.

The third primary standard is engagement with educators and society. Two of the three secondary standards under this primary standard are covered in the evaluation form; that is, engagement with parents and with professional learning societies. For the engagement with parents, it is clear in item 19 on the evaluation form that it asks about communication with parents. However, the item for engagement with professional learning societies is vaguely worded. On the evaluation form, item 18 asks directly about the teacher’s relationship with colleagues specifically, whereas the seventh secondary standard is about engagement with professional learning societies. The relationships with the parents were discussed indirectly in the interviews. Rana and Fahad talked about the fact that in the schools where they teach, principals discourage them from communicating directly with parents and they are asked to report any student to the school counselor, who will contact the parents. They both wonder how they are evaluated on communicating with parents while encouraged not to do so. Other teachers have had different experiences and were usually encouraged to communicate with parents frequently, but they do not have the time to do so, as expressed by Ali and Sara. Professional community engagement was covered in the survey (see Figure 19). Mixed responses were received; however, there seems to be an engagement of teachers with daily schoolwork and shared activities with other teachers. Yet, there was a semi-consensus in the interviews about not
having the time or the will to engage in extra activities with other teachers or professionals.

Fahad, for example, clearly expresses his annoyance with the continuous changes mandated by the Ministry in the regulations, guidelines, and rules. He thinks that teachers have had enough of this and that they are not willing to devote extra time to engaging with other teachers.

The third secondary standard in the third primary standard is about engagement with the local community. This standard is not reflected in the evaluation form. The survey data shows that about 50% of the respondents think their relationship with the community was of high importance in the latest evaluation they encountered. The interviews with teachers did not cover this standard directly, but it could be conceived from some responses that it is not something they are entitled to do. For example, while talking about recommendations, Sara mentioned that she would love to have an opportunity to engage with the local community and be part of activities and programs for them, which means she currently does not have this opportunity.

The fourth primary standard is being aware of linguistic and quantitative skills. Three of the five secondary standards here have to do with the appropriate use of the language. Instead of one general item measuring linguistic skills on the current evaluation form, there are three standards that are more specific and focused. More than 75% of the survey respondents think that their linguistic skills are moderate to high importance in the latest evaluation they encountered. This also aligns with the interviews, where five teachers have mentioned linguistic skills specifically during the discussion about the evaluation form. The other two secondary standards about numeracy and data collection are not covered in the evaluation form.

The fifth primary standard is about the knowledge of learners and how they learn. There are four secondary standards under this primary standard and two of them are not reflected in the evaluation form; that is, the child development phases and the characteristics of students with
special needs. The third secondary standard is concerned with individual differences and is partially represented in the evaluation form on item 7, which is a double-barreled item. The last secondary standard is concerned with the process of learning and is included in the evaluation form. The survey asked about the teachers’ ability to meet specific student needs, and about 85% of the respondents think this standard was of moderate to high importance on their last evaluations. Hind, in her interview, did mention that the principal and the superintendent do ask about the strategies she uses with some underperforming students. Rahaf also talks about the importance of having this standard, as it helps the teacher be aware of how the students are processing any new knowledge and information.

The sixth primary standard is about knowing the content of the subject the teacher teaches and knowing how to teach it. The two secondary standards under this primary standard are included in the evaluation form. More than 88% of the respondents on the survey think knowing the content of the subject taught was an aspect of the evaluation they went through. Teachers in the interviews were also in agreement that this is one of the most important aspects of evaluating a teacher. Hind, for example, states, “no teacher would ever argue about the importance of being knowledgeable about the content you are teaching.” However, five of the teachers have had an issue with the principals being the evaluators of this standard while they are not specialized in the same subject area. Fahad states:

It is unjust to have a principal who is from a social science background evaluating me in my level of knowledge in math or religion. The superintendents are responsible for evaluating that but most times they are not part of the evaluation, and the principal is the one who will do it.

The seventh primary standard is knowledge about general teaching methods. The two secondary standards under this primary standard are represented in the evaluation form. The survey asked the teachers about the extent to which the latest evaluation they encountered has
covered their knowledge of instructional practices. The vast majority think this was of high
importance. In the interviews, the teachers also endorsed this standard and thought that it is
crucial for the teacher to be knowledgeable of the general teaching methods.

The eighth primary standard is concerned with planning for lessons and being able to
execute the plan. The evaluation form covers two of the four secondary standards under this
primary standard; that is, planning for instruction and the use of different instructional methods.
The standards designing educational programs according to the teaching plan and linking shared
content across curriculums are both not covered in the form. Both the survey and the interviews
have not focused on these detailed aspects of instruction, but it could be argued that in general,
teachers do support having such standards, as it is essential for a teacher to be able to
demonstrate such knowledge and skills.

The ninth primary standard is the ability of the teacher to create a supportive and
interactive environment in the classroom. The evaluation form does not cover any of the
secondary standards under this primary standard. In the survey, the teachers were asked if their
ability to create a communicative learning environment was part of their latest evaluation. Again,
the majority thought this was considered; however, it is uncertain how this was measured, as the
form does not have any item covering this aspect. The interviews with teachers did not go into
details about the skills the teacher needs to master and be evaluated on.

The tenth and last primary standard is evaluation and assessment. Two of the secondary
standards under this primary standard are represented partially in the evaluation form; that is,
planning for the assessment and applying the assessment. The added standard is the utilization of
the assessment outcomes in future instruction. On the survey, the item about assessment
procedures received the highest endorsement from teachers as being part of the latest evaluation they have encountered. The interviews did not discuss these aspects of assessment in detail.

To sum up this section, the evaluation form that is used currently to evaluate teachers in Saudi Arabia does cover some of the STS. However, there are important aspects of the standards that are not covered in the instrument. The teachers’ perspective according to the interviews was not in favor of the current form, as it has not been updated in an exceptionally long time. It is also subjective, and principals vary to a great extent in the way they use and complete this form. To my knowledge, the instrument has not been validated in any previous study, as it has been difficult if not impossible to get to the data produced by these forms. Not only did the teachers have issues with the current evaluation form, but the key informants who were interviewed expressed that the form needs to be changed as well. In the next section, the focus is on what needs to be done to elevate the status of teacher evaluation and make it more relevant and aligned to the STS. The data comes from both the teacher and the key informant interviews.

**Teachers’ and Key Informants’ Perceptions on the Challenges and Recommendations for Teacher Evaluation**

In this section, the third research question of this dissertation is addressed; that is, what does the Saudi teacher evaluation system need to align with the STS and the best international practices? The comparison which took place in the previous section, the data collected from the open-ended questions on the teachers’ survey, the interviews with teachers, and the interviews with key informants are all used to answer this question. First, the challenges teachers and key informants have mentioned are presented. Then, the need to address the misalignment and the challenges is presented in three distinct categories. First, the needs according to best OECD practices are addressed. Second, the needs according to the teachers are addressed. Finally, the needs according to the key informants are addressed. In the discussion chapter, the researcher
broadly discusses all these needs and challenges and puts together a suggested framework for teacher evaluation. All the needs and recommendations discussed in this chapter are considered as findings based on the teachers and key informant responses. The researcher’s recommendations are discussed in the next chapter.

**Challenges Facing Teacher Evaluation in Saudi Arabia**

The thematic analysis of the teacher interviews and the open-ended question about struggles and challenges facing teacher evaluation revealed five main categories of challenges teachers think are negatively affecting the teacher evaluation system in the country. Each of these challenges was represented as a subtheme under the main theme “struggles and challenges.” These themes include teaching conditions, inconsistency, bureaucracy, system flaws, corruption, and nepotism. Obviously, the themes are connected and influence each other; however, the categorization was based on the teachers’ views. As the survey’s open-ended responses about struggles and challenges were added to this analysis, the number of codes in each subtheme is relatively higher than the codes in other themes. Quantification of some main ideas will first take place, then some examples of what teachers had to say will follow. Each of the subthemes are presented alongside examples of what the teachers had to say.

Most codes were found related to the corruption and nepotism subtheme. There were 62 codes under this subtheme. Words that are related to nepotism in Arabic such as (محسوبية/ محاباة/ تحيز/ مجاملة/ واسطة) appeared over 52 times. Words related to corruption such as (فساد/ مصداقية/ تساهل/ مزاجية) appeared 15 times. The most repeated idea by most of the teachers is that principals will subjectively evaluate teachers based on how they feel about them and if they have a good personal relationship with them or not. Another idea that was frequently raised by many teachers is that evaluations are used as a courtesy by principals for those with whom they have a good
relationship. Saad, for example, states that he worked under 18 principals and most of them treat teacher evaluation as a courtesy. If you are in a good relationship with them, you should never worry about the evaluation. Rahaf is another example where relationships with the principal caused her to get a relatively lower score on her evaluation. She states that she was new at the school and wanted to keep her relationship with the principal formal and do what she is obligated to do without giving any chance of extra work, as she was having family issues.

The second most coded subtheme is system flaws. There were about 42 codes under this subtheme, where teachers describe aspects of the teacher evaluation system they think affect its utility. Criticizing evaluators was the most frequent code when talking about challenges. In about 15 codes, teachers described how principals lack the minimum acceptable level of handling the evaluation of teachers. Ali, who was a principal for about 20 years, describes how he was not given any support or training on how to evaluate teachers after he became a principal. The second challenge that was mentioned by teachers is the lack of any consequences for teacher evaluation. There were about 10 codes related to consequences, with the majority focusing on the lack of incentives. One teacher from the survey respondents stated, “Why should I care about being evaluated when I know I will get the same score as other lazy teachers and will not receive any incentives? There is no motivation at all to work hard.”

The third subtheme, under the theme challenges, is teachers’ working conditions. There were 36 codes under this subtheme. The most frequent challenge here was classroom size. About 20 teachers mentioned that class sizes with 40 or more students will never enable teachers to bring the best they have. A second challenge was the stress teachers face in terms of the number of tasks and obligations they must do. Sara mentions that above the 24 lessons per week she must teach, she is also responsible for many extracurricular activities and programs. She must also
share supervising during recess, student drop-off, and pick-up. Another challenge raised by participants and interviewees was the lack of facilities and resources at many of the schools. Lack of facilities and resources in schools was mentioned by about 10 teachers. The status and promotion of teachers in society was also a challenge, according to seven teachers. The least challenging aspect that was mentioned is the financial status of teachers, which indicates this is not generally considered a struggle for teachers in the country.

The subthemes inconsistency and bureaucracy come next at the same level with 18 codes for each. In the subtheme inconsistency, teachers mentioned the discrepancy they face in the procedures of their evaluations from school to school and from principal to principal. According to Arwa, some principals do take the evaluation seriously enough, but the system does not help them to be honest and serious, as there are no consequences for the evaluation. Others will just treat these evaluations as routine office work that could be done with minimum effort. Sara as well talked about the discrepancy in the school’s facilities and principals’ capabilities. She touched on the idea that a centralized system does not work because of these discrepancies.

**Recommendations in Accordance with OECD Best Practices**

The OECD presents a valuable comparative data resource between countries on various levels. Considering teacher evaluation, the OECD has analyzed and studied the performance of the participating countries on several measures and came out with a list of practices that are critical to any teacher evaluation system. These practices are presented and discussed in the OECD report about effective teacher policies (Avvisati, 2018; Nusche, 2013). In this subsection, the researcher discusses whether these practices are evident in the Saudi teacher evaluation system or not based on the data collected from the survey and interviews.
The first suggestion by the OECD is establishing teaching standards to guide the evaluation of teachers and professional development. This suggestion has been established by the ETEC through the creation of the STS. These standards currently play a significant role in the licensure program for teachers; however, as this dissertation hypothesizes, is not yet implemented in the evaluation of teachers nor the professional development programs. Teachers in the survey and the interviews have talked about the evaluation form that is not completely aligned with the STS, as presented previously.

The second suggestion the OECD endorses is to resolve the tension between the developmental and accountability functions of teacher evaluation. According to the survey results, most of the teachers feel that the evaluation they go through currently is summative. The formative developmental aspect of the system is not obvious, to say the least. It may also not be aimed at by the system, as it is hard to fill. Fahad, in his interview, described how overwhelming the constant change in policies is, and how teachers and school administrators lack autonomy over the specifics of teacher evaluation. He states:

You cannot imagine the number of circulars we receive from the ministry about the specifics of teacher evaluation. Each time the principal is conducting an evaluation, he must go back to all these circulars and check what he should do. I wish there were a comprehensive guide for teacher evaluation that shows everything we need to know in one place.

Majed, a key informant who was interviewed, indicated that there has been a lack of alignment between teacher evaluation and teacher professional development for a long time and that currently, this misalignment extends to the teaching professional standards. He states:

It should be a main priority for the Ministry to develop a comprehensive system for the teaching profession which starts from the teacher preparation programs and covers the professional development teachers receive and the evaluation of teachers as an indicator of how the Ministry is doing in terms of developing and enhancing the teaching profession.
This shows that the aim of the teacher evaluation system is still not clear and needs to be specified. The system should address both the formative and the summative functions, with clear guidelines and procedures.

The OECD also suggests conducting multiple evaluations for the developmental aim. The occurrence of the evaluations in the Saudi teacher evaluation system is evident; however, this could not be said that it is for a developmental aim. The recommendation also stresses the importance of utilizing multiple sources of evidence in the evaluations. This is clearly not the case in Saudi Arabia, where teacher evaluation is completely controlled by the school principal through a classroom observation that may or may not be conducted, as discussed previously. The same recommendation also states that the evaluators responsible for conducting the evaluated teachers should be competent and well trained, as an elevated level of competence helps in making the evaluations more valid and reliable. The teachers in the interviews have signaled in multiple places that the principals, who are the main evaluators, are not competent enough to be able to do that task. Ali, a previous principal, has stated clearly that after he was appointed as a principal, he was asked to conduct evaluations without any guidance or training.

Another important recommendation by the OECD is that the evaluation system should feed into professional and school development. The results from the interview and the survey indicate that teacher evaluation works in isolation. Nora, when asked about the consequences of the evaluation and how the results are used, clearly responded with “nothing happens.” Hind echoed that by describing how the form is not even given to the teacher unless she pleads for them, and that the results are archived in their hard copies and not entered into software.

Having external evaluators who periodically evaluate teachers is an important recommendation by the OECD. The Saudi teacher evaluation system has the superintendents as
external evaluators. Although they might work with the same schools for a long time and do have effective communication with the principal and the teachers in the school, the evaluations done by the superintendents are usually not final until the school principal approves them. Basil, a key informant, described the fact that all superintendents are in fact internal evaluators, as they previously have been principals and/or teachers at the same schools they now oversee. He also thinks that this is one of the main reasons the system is filled with nepotism and corruption.

The last recommendation for the OECD is about the preparations that should be done for all teachers before the evaluation. The teacher should be thoroughly informed of the policies, processes, and consequences of teacher evaluation. For the summative evaluations that are done by external evaluators, the teacher should also be fully aware of the appeal process and how to discuss the results of the evaluation and make use of them. It has been discussed previously that the teacher evaluation policies in Saudi Arabia are not well established and documented, which leaves the teachers not fully aware of the system. Sara talked about how she was surprised by an unexpected visit when she was not ready to be evaluated, and that this was the only time she was evaluated during the whole semester. She states, “It would have been okay with me if the evaluation was to provide feedback and not to judge my professionalism.” She refers to the evaluation as being summative and judgmental and occurring once every semester; hence, she wanted to be notified before it took place.

Recommendations by Teachers

In this subsection, the focus is on the recommendations given by the teachers in both the survey open-ended questions and the interviews. Recommendations are listed as they were presented by the teachers (the discussion of all the needs will take place in the Discussion chapter). The theme recommendations had 129 codes from survey participants and 41 codes
from the teacher interviews. Therefore, I had to create subthemes to categorize these recommendations and make them more specific. Six subthemes were created. These themes are autonomy, working conditions, professional development, methods, evaluators, and incentives.

The subtheme with the most codes (i.e., 37 codes) was evaluators. There were 30 teachers expressing how crucial it is for the evaluation system to have evaluators other than or in addition to school principals. Many talked about the fact that school principals often misuse the power of teacher evaluation against teachers. By adding more evaluators, either external or internal, where the final score is distributed and not controlled by the principal, the validity and honesty of the evaluation system will dramatically increase. About 10 teachers have recommended having an internal superintendent in the school who is responsible for conducting periodic developmental evaluations, so the teacher is ready when the external judgmental evaluator is visiting. Doing such will enable the system to be formative and decisive at the same time, as noted by Fahad. Other teachers have encouraged having parents, students, and peer teachers as evaluators. Others have also included self-evaluation as a means for raising teacher self-awareness and self-development.

The second subtheme under the recommendation theme was labeled incentives. There were 27 codes under this subtheme, where 24 teachers describe how important it is to have a clear incentive scheme that motivates teachers to bring out the best of their skills and knowledge. Hind talked about how crucial it is for an evaluation system to discriminate between teachers with high performance and those with low performance. She states, “Where is the motivation when I know that I will get the same score, if not lower, as other teachers just because they are relatives or friends with the principal? It would not make any sense to work hard.” About 20 out of the 24 teachers specified material incentives are what the system needs, whereas about seven
teachers indicated that a just and discriminating evaluation system is enough of an incentive. What they mean by that is that being recognized as a high-performing teacher by a robust and honest evaluation system has an intrinsic value.

The third subtheme was related to the working conditions of teachers. There were 25 codes under this subtheme. The most frequent recommendation is to have smaller classrooms and less teaching load. A respondent to the survey states:

How can I make sure my students are progressing academically when I have more than 40 students in the classroom, and I teach 25 lessons per week? It would make no sense to judge me based on the students’ achievement, I don't think I am a super teacher.

The subtheme professional development comes next, with 20 codes. Teachers were asking for more professional development that is relevant to their work in the classroom and in line with what they need. They also mentioned the importance of professional development for principals and superintendents who are responsible for evaluating teachers. Hamad, for example, talked about how unfortunate it is to have principals who are appointed without any professional training on how to lead schools and teachers. He states, “Due to the need, you could be a teacher today and a principal tomorrow without any training or preparation.”

Several evaluation methods were recommended by teachers in the next subtheme. The focus of these recommendations was on the importance of having more than one visit, as it is not enough to judge a teacher’s performance based on one short visit. Khalid suggested that the evaluation should be done monthly to serve the developmental aspect of evaluation. Five teachers were in favor of giving the teacher a chance to evaluate himself. Hind argues that this would be a meaningful change for the system and would transfer the responsibility for development to the teacher. A couple of teachers recommended having an electronic portfolio of all the teacher’s achievements that could also be part of the evaluation process.
The last subtheme under the recommendations theme is teacher autonomy. Twelve teachers have noted the need for a system that gives the teachers some freedom to decide on what is better for the students. Hamad argues that the current system views teachers as machines rather than professionals. He states, “The system is not in favor of a creative teacher, as there are no incentives or freedom to decide on anything in the teaching and learning processes. That’s why most teachers are just doing the minimum requirements.” Three teachers also mentioned the importance of restoring the image of teaching and teachers, which has suffered over the past years. One of the survey respondents argues that the teacher’s image plays a significant role in the way he or she acts and performs in the classroom. Another teacher from the survey respondents states:

The centralized system takes all the autonomy from the teachers in their classrooms and the principals in their schools and puts it in the hands of a bureaucratic system. Give us the freedom to decide on what fits our schools and classrooms and then evaluate the outputs.

Recommendations by Key Informants

In this subsection, the results of the key informants’ interviews are presented. The analysis consisted of two main aspects, the struggles, and the needs. In the struggles subtheme, the key informants echoed to a great extent what the teachers had to say about the challenges and struggles facing teacher evaluation. The struggles and challenges start from the fundamentals of the current system, says Majed. There is not any alignment between any of the aspects related to teacher evaluation. He states, “The practice of teacher evaluation is looked at as an administrative burden that principals just want to be done with. It is not connected by any means to any developmental or accountability aspects.” Basil also referred to the entire system as being built with little effort and without considering any academic or professional guidance. He states,
“No one to this day knows why exactly the Ministry evaluates teachers, as there is no clear guide to the principals or teachers or the education community of how the results are used or utilized.”

Another struggle that was brought up by the key informants is the lack of objectivity in the system. They all agreed that putting power solely in the hands of principals made it a tool to implicitly threaten and manipulate teachers in some cases and to praise and honor teachers who are close to the principal in others. Adel states, “Our culture is based on complement and praise. It is why we see the evaluation scores so inflated and do not discriminate between good and bad teachers.” Basil as well noted that the system is controlled by principals who are not prepared or trained to evaluate teachers. The key informants also discussed the idea that the current form that is used to evaluate teachers does not have any scientific basis to it. All three interviewees admitted that they have never seen a study done on the validity or reliability of this tool. Adel took it even further and stated that any data delivered by this tool lacks the minimum requirement of valid or suitable data to be part of a scientific study. Lastly, the key informants indirectly referred to the bureaucracy of the system and the unstable relationship between different government bodies overseeing education in the country and how it resulted in a gap between theory and practice.

In terms of the needs, the key informants expressed a range of developments they think need to take place for the teacher evaluation system to be fair and beneficial. First, is the importance of cooperation among responsible bodies in the government so all the initiatives are coordinated. Basil covered that issue in detail as he used to work at the Ministry of Education. He argues that there have been so many developments and innovative ideas in the Ministry, but since there is no coordination with other players in the education field, these ideas never get accomplished. He also referred to the challenge the Ministry itself faces in terms of its
bureaucratic way of handling things. Projects may take years to get approved and different departments have their own ways and pace of handling things, and new projects should be looked at in each of these departments. Majed clearly states the importance of a valid evaluation system in that it aids in identifying the gap between the ideal professional practice (i.e., the teaching standards) and actual practice so educators and policymakers can bridge this gap. Adel stresses the importance of having a comprehensive development in the entire system. He argues that this is the only way to have a consistent and useful improvement. He states:

The entire system should be included in an overhaul. Gains from having a proper teacher evaluation system will be limited if we do not develop the teacher education programs to align with the teaching standards, for example. When underqualified teachers enter the teaching profession, they will require extensive resources to get them to a higher level of proficiency. Therefore, we should include all aspects of teacher preparation, professional development, and evaluation in any plan.

Another important aspect from the key informants’ perspective is to adopt a system that is based on evidence. Basil argues that it is crucial to have an electronic platform where teachers can upload evidence of their performance that gets reviewed by the internal evaluator and the principal before having an external evaluator look at it. Professional development, he stresses, should also be based on reviews from this platform, the teacher performance on the licensure tests, and the periodical classroom visits.

**Conclusion**

In this chapter, the aim was to present the findings utilizing different lenses and analyses from different data sets about teacher evaluation. These findings helped in answering the research questions and building a comprehensive picture of teacher evaluation in Saudi Arabia.

In the first section, the question about the current teacher evaluation system in Saudi Arabia was answered. The teachers’ lens was used to address this question. Overall, the current
state of teacher evaluation in Saudi Arabia is not encouraging. Teachers have reported that they are not aware of the purpose of the system and that the procedures that take place within the system are subjective and continuously changing. Classroom observation is commonly reported as the only data source for teacher evaluation, and school principals are the main evaluator of teachers. Professional development is highly selective by school principals, and it is not linked in any way to the results of teacher evaluation.

In the second section, the Saudi teacher evaluation system was compared to the OECD best practices and recommendations on teacher evaluation. This also contributed to answering the first research question, which was concerned with the comparison between the Saudi teacher evaluation system and international best practices. The lenses that were utilized for this comparison were best practices and the leading teacher evaluation system. The comparison was based on four main indicators, which are: (1) the existence of a framework for teacher evaluation, (2) the procedures of teacher evaluation, (3) the evaluators, and (4) the consequences of the evaluations. The Saudi teacher evaluation system varied in its compliance with the recommended practices by the OECD. For example, the system does have a framework for the teaching professional standards, yet, is falling behind in utilizing different data sources to evaluate teachers. On the consequences of teacher evaluation, the Saudi system is not benefiting in any way from the data collected on teacher performance, as there is no plan or system to do so and the data that is collected lacks the minimum requirements to be valid and reliable data.

The third section discussed the alignment between the national teaching professional standards and the teacher evaluation system in place, which is the core of this dissertation. After looking at the teacher responses, interviews, and the observation form that is used in evaluating teachers, it could be argued that the evaluation practices do align to some extent with the STS;
however, there are standards that are not yet covered. The current practice is more of a subjective, general view on how the teacher is performing, whereas the STS focuses on specific skills and expertise teachers must demonstrate.

The last section addressed the challenges facing the teacher evaluation system and the recommendations to overcome them. Data from teachers and key informants were two key sources of evidence utilized in the analyses and discussion in this section. The main ideas discussed here were the importance of diversifying the evaluators, a need for more focus on incentives, improving working conditions by decreasing the number of students in classrooms and the workload, developing more practical and applicable professional development programs for teachers, proper training of the evaluators, increasing teacher autonomy, cooperation between different departments within the Ministry, and more research and scientific based decisions. It could be argued that the entire system needs to be reorganized and linked effectively to other domains of teacher preparation and development for it to be fruitful. The adaptation of multiple sources of data to evaluate teachers and the distribution of the evaluation power amongst different stakeholders are essential suggestions for the system. These were the main recommendations by the study’s participants. They were listed and discussed in this chapter as a response to the third research question. In the next chapter, a discussion about all the findings will take place alongside a proposed framework for teacher evaluation and a list of recommendation that would arise based on the synthesis of the findings of this research.
CHAPTER V
DISCUSSION

The main aim of this dissertation was to investigate the status of teacher evaluation in Saudi Arabia utilizing several lenses to interpret that data that would aid in precise recommendations to help elevate its status. Specifically, these lenses aided in providing the findings that answered the research questions in the previous chapter. In the current chapter, the focus is directed towards the researcher’s interpretations, insights, and recommendations. The chapter starts with an overview of the dissertation. Then, a set of recommendations covers various aspects of teacher evaluation. The chapter ends with a discussion of the significance of the research and its limitations.

Overview of the Dissertation

To reach the main aim of this dissertation, the current teacher evaluation system had to be studied through the perspective of teachers and key informants in the Ministry of Education. An explanatory sequential mixed design was followed to allow the researcher to depict a broad, yet detailed picture of the current evaluation system and how it relates to the Saudi Teaching Standards (STS). A survey, completed by 642 teachers, was followed with two sets of semi-structured interviews, one with 11 teachers, and the other with three key informants. Each of these steps was informed by the initial results of the previous steps. Quantitative data collected from the survey was entered into SPSS to check for its reliability and factor structure before conducting descriptive and inferential analyses. The qualitative data was loaded into MAXQDA to perform a thematic analysis on all the interviews.

The results of the study indicate that teacher evaluation in Saudi Arabia has been static for over 3 decades. There is no clear guidance or standards on what the purpose of evaluating
teachers is, how it should be done, and what are its consequences. This ambiguity has led the practice to vary enormously from school to school and from principal to principal. In the previous chapter, a rather complete description of the practice of teacher evaluation in Saudi Arabia was provided. In short, teachers expect to be visited by the superintendent or the school principal at least once a year. This is more consistent for novice teachers compared to more experienced teachers. Principals are the only evaluators whose evaluations count, and they generally have full authority over the evaluation process. The observation form that is used during the classroom visits remained the same for an exceptionally long time and it has not been validated due to not having valid data that could be used for this purpose. As for the teachers’ understandings of the evaluation system, most of them think it leans heavily towards being summative more than formative. What makes them think this way is the lack of feedback engagement and that professional development is not required. According to the teachers, the scores that result from the evaluations are highly subjective and meaningless.

The evaluation form does cover a great amount of the STS; however, there are some standards that are not covered. The least covered standards were the amount of professional development the teacher had to complete and promotion of cultural diversity in the classroom. It is important to note that what this study intended to find is the content coverage of the STS in the practice of evaluating teachers. It did not, however, aim to find out if the current observation form is a valid measure of these standards. The key informants have illustrated how the teacher evaluation system is out of date and urgently needs to be updated to align with the standards to raise the quality and efficiency of the education system in the country.

Compared to international OECD teacher evaluation practices, the Saudi system lacks a comprehensive framework for teacher evaluation, which is a fundamental aspect of teacher
evaluation in most of the OECD countries. Classroom observations are the sole method of collecting data about Saudi teachers, and it is the most common method used in OECD countries as well. However, it is advised to have more than one method of collecting data for evaluating teachers, which the Saudi system lacks. Also, the complete dependency on school principals in approving the evaluation report and score without any consideration given to input from other stakeholders undermines the validity and objectivity of the evaluation process and does not align with the widespread practice in OECD countries with high-ranking education systems. Lastly, the consequences of the evaluation system in Saudi Arabia seems to be missing. The widespread practice in leading OECD countries in education is to have a system of evaluation that feeds into the development of teachers and schools in addition to other summative consequences for individual teachers. However, the Saudi teacher evaluation system is not utilizing the results of the evaluations in any way. This could be a result of a lack of confidence in the validity and reliability of such system.

During the process of the research, the researcher has made all the effort to maintain a highly valid and reliable process. For the quantitative part of the study, the survey was revised through a cognitive interview process with three teachers to ensure all questions and words used in the survey are understood as intended by the researcher. A pilot administration was also done to ensure the appropriateness of all questions. A factor analysis for the whole survey yielded four factors, as anticipated and the Cronbach alpha was calculated for each factor separately and for the whole survey to check its reliability. All these coefficients revealed high reliabilities. In Chapter III, a detailed discussion about the process of validation was presented.

For the qualitative part of the study, the interview transcripts were sent to the participants to make sure the ideas and discussions were captured correctly. This member checking helps in
establishing the internal validity of the study. It was interesting to find out how the discussions in the interviews aligned with the survey results frequently, but it would also contradict what have been found from the survey in some instances, which enabled the researcher to collect different perspectives of a complex issue. The several strands in the mixed methods design allowed for confirmability of the reached results and strengthened its generalizability. The various stages of the study also allowed for the results to be triangulated and tested. Member checking also allowed for the interpretations and conclusions to be validated by the participants. It is important to note, nevertheless, that the transformability and generalizability of the findings of this study are bounded with the limitations of the current study and its complexity.

All human produced literature is bounded and limited (Simon & Goes, 2013). For this study, the complexity of the issue in hand makes it impossible to cover all its aspects and dimensions. Therefore, the focus was on the teachers’ and key informants’ perspectives. Utilizing a mixed method design has also strengthened the trustworthiness and generalizability of the study findings. Teachers were fairly represented through both strands of the design, the survey and the interviews. However, it was unfortunate to not be able to interview key informants who are working on teacher evaluation policies presently. It is understood that personal holding such positions may feel a bit sensitive and conservative when it comes to sharing their thoughts and plans. However, it would be so beneficial to the research community and the development of any field to see more engagement between researchers and policymakers to incorporate what scientists and experts think in any new policies. It will also aid in moving closer to decisions and policies that are based on evidence. Another major limitation to this whole study is that it was conducted during the COVID-19 pandemic. All processes, stages, and meetings were held virtually as a result.
Recommendations and Policy Options for Teacher Evaluation in Saudi Arabia

Any evaluation system should start with a solid knowledge base to aid in bringing together all its elements and provide a coherent presentation about its procedures and expectations. The Saudi national standards play a vital role in setting what it expected from teachers, but there seems to be a need for an overall framework for teacher evaluation. A recommended framework that is based on what is the best practice of teacher evaluation and what would help develop teacher evaluation in Saudi Arabia is put forward to provide an overview of the systematic functions of teacher evaluation. Figure 20 depicts the most important aspects of any teacher evaluation system. Such a depiction makes it easier to trace the various aspects of the system logically and to facilitate the comprehension of the system, its various components, and how they are linked to each other. It also opens the doors to build the necessary tools to measure the performance of each of these components to reach the desired impact of the system. This system should also be part of a bigger picture depicting the vision of the whole education system, starting with teacher preparation programs, as they also have an extreme impact on teacher effectiveness.

The suggested framework, which is adapted from a framework developed by the OECD (2013) starts with the inputs to the evaluation system, which consist of government and system contributions and resources. The governance inputs include all legislative aspects of the system such as the framework under which teacher evaluation works, the determination of stakeholder groups and their respective responsibilities, the objectives of the evaluation system and its consequences, and lastly the determination of the teaching standards in which the performance of teachers are evaluated. The capacity inputs include aspects within the system that would affect its performance. Evaluators’ familiarity with the system, the level of training they received, and
the skills they hold are examples of system capacities that need to be enlarged and measured. Another aspect of the system capacity is the level of effectiveness teachers already have. This is partially a result of the teacher preparation programs and the previous teaching experience they have had in the past. The available tools for the evaluations such as the observation forms and any required software that would increase the efficiency of the evaluation process, is also an important capacity aspect.

The evaluation procedures in the middle of the figure represent the methods used to collect data about the performance of the teacher. It is recommended for sources to be multiple (Schleicher, 2020), as the more data sources the system collects, the more the results are reliable and valid. The utility of the evaluation depicts the main function it plays in the education system. If a summative system is being considered, then the results will have to consider the career progression, licensure programs, and annual raises and promotions. If the system is designed to be developmental, then the feedback loop, the professional development opportunities, and the professional community enrichment should be considered. While aiming for both evaluation types is challenging and the focus may end up shifting to one of them, it is still recommended to try and add a developmental aspect to any teacher evaluation system (Firestone & Donaldson, 2019). The overall intended outcome is to improve the effectiveness of teachers, which will likely result in higher education quality. It is crucial for the system to collect data that could be used later to evaluate the entire process.
The teacher evaluation system is a "wicked" one that would require patience and consistency to develop. The overall development of teacher evaluation should recognize developing the whole system including all aspects that affects and get affected by it. Without such a systematic approach to its development, all efforts to improve teacher evaluation would not pay off. The overlapping and connectivity between teacher evaluation and professional development, for instance, makes it hard to get tangible development without reforming both simultaneously. Another important factor in the development that one could infer from the suggested model is that a development of such a large-scale system should always start by piloting parts of the interventions on small rural schools to test for its efficiency and
effectiveness, and to make sure there are no unintended consequences for these changes. The outcomes of these developments should also be assessed and evaluated on different levels before rolling out the reform. In the next section I will be discussing some of the policy options which I think are necessary to develop a more affective teacher evaluation system based on the results of this study.

**Teacher Evaluation Policy Options**

Based on the framework above, several policy and practical recommendations are discussed below. It is important to note that these recommendations are not expected to properly work if applied in isolation or only for a limited time span. Instead, these recommendations should be institutionalized into the system so all stakeholders make them part of their daily job (Stufflebeam, 2002). This, however, may be extremely hard to achieve due to the complexity of such a system.

It may also be beneficial to revisit the main recommendations by the teachers and the key informants in the previous chapter before delving in the researcher's recommendations. Most of the interviewees focused on the practices of teacher evaluation and the importance of diversifying the data sources about the teachers’ performance. The cooperation between different departments within the Ministry was a key recommendation of the key informants. Teachers also made several suggestions. First, the teachers suggested some development in the working conditions of teachers such as decreasing the classroom sizes and the workload. Professional development that is related to teachers’ everyday work was also proposed by the teachers. Lastly, the teachers also stressed on the importance of training the evaluators so they can perform professional evaluations. These, and other findings, were considered and synthesized with the
available literature on teacher evaluation to come up with the below recommendations and policy options.

**A Hint of Decentralization**

Policy changes may have a huge effect on how effective teacher evaluation is in Saudi Arabia. In this sub-section, policy options that are based on the findings of the study are presented and discussed. First, the Saudi education system is centralized, where all aspects of school governance are controlled by the MoE. The control on schools the Ministry maintains has its own benefits to the education system such as enabling widespread regulated reform that based on a national framework and standards. For example, the STS document that was issued by the Education and Training Evaluation Commission (ETEC), an education regulatory body, was extremely beneficial in standardizing the teaching profession across the country. This standardization plays a crucial role as an evaluation criterion for teachers. However, on the other hand, excessive centralization diminishes the level of autonomy schools and teachers hold (Mead et al., 2012). School autonomy and teacher autonomy are both crucial factors to consider while developing an evaluation system. Holding entities responsible, which is a major aim for evaluation, should also be accompanied with the confidence and a margin of freedom to decide on issues that occur in the school or teacher community (Nusche, 2013, p. 283).

The current study has not investigated the different models of school leadership and governance but the recommendations based on PISA and TALIS results are clear on how important it is for teachers and schools to gain some control on how to navigate everyday issues and how to increase the effectiveness levels of teachers (OECD, 2019c). Darling-Hammond (2012) discussed the importance of having some level of autonomy for schools to develop, for teachers to grow, and for teacher professional communities to flourish. It provides them with
opportunity to navigate what works better for them considering the differences in localities and needs across a vast country. A hint of decentralization will not only provide the schools and teachers with some level of autonomy, but it will also strengthen their commitment and loyalty to their profession and motivate them to be an active part of the system.

**Clarity**

Communication between the centralized authority of education and the teachers seems to be a challenge in Saudi Arabia, according to the teachers who participated in this study. Many complained about the ongoing changes and orders they receive from the Ministry. In some cases, there are new guides and orders every time teachers are evaluated. It is not clear if these changes and orders are part of a bigger development plan but there were no indications of such plan during the past years. Another finding was that a sizable portion of teachers were not informed of the STS and some cases they have shown resistance to even look at the standards. It is crucial that the entire system is fairly informed about the framework behind the system and its objectives and methods. This will ensure the most crucial stakeholder of the system is well informed, as most development plans fail to succeed due the lack of transparency and clarity on the side of the implementers, in this case the teachers.

**A Developmental Focus**

It is recommended for the evaluation system to merge the two main aims of teacher evaluation, the developmental and the summative (Darling-Hammond, 2012; Nusche, 2013). However, that is a challenging task (Popham, 2013), as evaluation systems tend to lean towards one objective over the other. The result of each type of evaluation is different, as are the methods and data collected for each objective. Developmental evaluation results intend to fix teachers
instead of firing them. It is thus useful for the system to differentiate between the methods and
data serving the developmental aim from those serving a judgmental aim.

Darling-Hammond et al. (2017) discussed how developmental formative teacher
evaluations are more capable of sustaining the growth and development of teacher communities.
Teachers are usually less opposing to such policies and procedures, as they aim to engage them
in an ongoing development of their profession. This is unlike judgmental evaluations, where
teachers’ performance is judged and high-stake decisions are made. These high-stake decisions
are useful and crucial in many cases, but they require a higher level of validation of the tools and
the data it uses. Overall, the main challenge is not to produce new methods to evaluate teachers,
but rather to make sure these evaluation methods are sound, supportive, and do not intend to do
harm where harm is not warranted.

From the findings of this study, it appears that teacher evaluation in Saudi Arabia is not
providing enough, if any, developmental and supportive elements to the teachers. The
professional development is not linked to the performance evaluation of teachers, and evaluation
results are, for the most part, not utilized in any way. Based on these findings and the literature, a
developmental focused evaluation done by internal evaluators alongside periodic summative
visits by external evaluators may help in addressing both evaluation types. It is crucial, however,
to differentiate between the methods used for each.

**Evaluation for Teacher Benefit**

The previous three recommendations, a degree of decentralization, clarity and
transparency, and the developmental focus all stem from the importance of teachers' perceptions
and understanding that evaluations should be for their benefit instead of being against them.
They should be able to feel the support and encouragement from the MoE in all evaluation
documentations, from internal and external evaluators, and from the professional training that is offered. Teachers should also have a voice and be engaged in discussions about improving teacher evaluations; this can be done through interviews, focus groups, and surveys. The experience of interviewing teachers in this study has shown the importance of giving teachers space to express their thoughts and experiences. Engaging teachers in the process of improving teacher evaluations will help institutionalize evaluation practices and help make them part of the teaching culture.

**Professional Learning Communities**

Professional learning communities are also crucial in the institutionalizing process of evaluation. Learning is more effective when it is done collectively in homogeneous groups sharing the same aim and goal (Voelkel & Chrispeels, 2017). Communities of teachers are usually more aware of the shortcomings they face in their everyday practice than any policymaker who rarely visits schools, and they should collectively work to bridge any gaps. It is amongst the most important steps in the process of education development to build this type of organizational learning capacity in our schools and teachers, and to give them the trust and autonomy they need to be proactive in creating the sustainable change that is desired for education. To achieve effective professional learning communities, there are several elements that should be considered and deemed according to Vescio et al. (2008). First, a shared vision and beliefs about aspects of teaching and learning is required. These beliefs are assumed to be developed throughout teacher preparation programs prior to becoming a teacher. Examples of these beliefs is the role of teachers and schools in the preparation of future generations, and the importance of knowing how students learn and progress. Second, an unobstructed vision that raising students’ learning experiences and outcomes is the shared goal for teachers in
professional learning communities is also beneficial. Third, a critical reflective environment is crucial in pushing these communities forward in terms of professional performance. In addition to these characteristics, the nature of these professional community groups requires the system not to be privatized, at least not entirely, as privatization may lessen the amount of flexibility and experience shared between teachers and schools. More experienced teachers should also be motivated and freed to mentor, guide, and support novice teachers in their first couple years in the profession.

As shown in the findings (Chapter IV), the level of collaboration and support between teachers in Saudi schools is low. Many teachers expressed how rare it is for them to meet with other teachers and talk about their experiences and challenges to try to elevate one another. The adaptation of professional learning communities would require major restructuring to the entire system, if done properly. Teachers and schools should have the ability to change the way they teach, what they focus on, and what professional development they need. These, and many other practices, are not currently evident in the Saudi system and will require a radical change in the entire system. In other countries, there is clear empirical evidence that demonstrates how impactful and beneficial professional learning communities would be for the education system (Vescio et al., 2008). A successful implementation of professional learning communities in Saudi Arabia should result in teachers having greater ownership of the evaluation system and feeling responsibility towards their profession.

**Evaluation Consequences**

Evaluation without consequences can be considered a waste of time and effort. It is crucial that teachers know and feel the consequences of the evaluation system, if the aim is to institutionalize teacher evaluations. Unfortunately, after deep exploration into this topic with
interviewees and the survey data, it is evident that in the Saudi context at the moment, teacher evaluations do not have any consequences, with one exception which will be discussed later. From a development orientation, teachers rarely get a chance to discuss their evaluations with the evaluator (i.e., the school principal). There are no actions taken based on the evaluation report or the score a teacher gets. Professional development programs are usually offered periodically regardless of the real need of teachers. In many cases, teachers do not receive their evaluation reports until the end of the academic year.

From the summative evaluation orientation, there is not that much of a consequence. Teaching in public schools in Saudi Arabia is considered a public service job; thus, it is unlikely for any teacher to be dismissed unless a serious felony is committed, but not for being an ineffective teacher. The only summative use of the evaluation scores, as reported by several teachers, is using it as a criterion for prioritizing teachers in the annual teacher transfer plan, where teachers who want to move to another city or province will have to collect points and the more points they collect the more likely they will be granted a transfer. However, this use had an unintended consequence. Principals started to give high scores for underperforming teachers to get them moved from their schools. It should always be in the mind of the policymakers that the group of stakeholders who are going to execute the evaluation plan are likely to manipulate the system in their interests.

Another main and crucial consequence any evaluation system should plan for is its ability to produce meaningful high-quality data that could be used to evaluate the system and for future research about teachers’ effectiveness and professional development. During this study, the researcher has not found any studies that were conducted using the data gathered through the evaluation of teachers in Saudi Arabia. This data is crucial for evidence-based decisions to be
made about teachers, their students, the education system as a whole. The recent vision and efforts articulated by the MoE are promising and need to be followed through with an inclusive and thoughtful approach.

**Diverse Trained Evaluators**

Evaluators are the heart of any evaluation system; therefore, they should always be skilled and professional. In the Saudi teacher evaluation system, principals are the primary, and in many cases the sole, evaluators. The power and authority they hold over teachers is immense and with the high subjectivity in the evaluation process and reports, the entire process is susceptible to manipulation and corruption. To avoid such abuse of power, principals should be trained rigorously on why and how to evaluate teachers, including how to make the process beneficial for the teacher and the school. Also, the evaluation authority and power should be distributed amongst more than one evaluator. Peer teachers, especially the tenured and more experienced teachers, should contribute to the evaluation of their colleagues. External evaluators and inspectors should also be part of an annual visit to schools to take part in the evaluation. Each of these stakeholders should have a pre-defined role and objective; however, their combined efforts should lead to one comprehensive and objective evaluation of the performance of teachers. A recommended way of distributing these roles is to assign formative evaluations to principals and to motivate them to do so through setting annual goals related to the teacher and school performance. To validate achieving these goals, external evaluators would visit the school once a year to review the evaluation process and conduct their own summative oriented evaluation. This way the collected data and reports made by the principals could be validated and checked for more objective and precise findings.
In addition to engaging more evaluators, they also should be trained not once but repeatedly on the purpose and procedures of evaluation. These training sessions should also be constructive, engaging, and open to discussion so evaluators can collectively work on developing the status of teacher evaluation. Because of turnover and because of the need for deeper learning, implementing the training just once would not be beneficial in terms of efforts to institutionalize teacher evaluation.

More Stakeholder Engagement

Teacher evaluation systems are usually complex due to the number of stakeholders who are engaged in teacher evaluations or benefit from them. Teachers, principals, students, parents, and policymakers are amongst the most important, and the society could also be part of the evaluation system, as teaching is considered a public service. Admittedly, it would be challenging to involve all stakeholders in the planning and practice of teacher evaluation; however, it is still important to include as many as possible. The more stakeholders a system includes, the more comprehensive and mature the system will be. In the Saudi context, it appears from the interviews and the survey data that the planning and policy creation for the teacher evaluation system is solely done by policymakers in the MoE. There might be some workshops for teachers, superintendents, or principals but the contributions of those are individual and do not represent the whole body of teachers or their voices. In the execution part of the evaluation, a diverse range of evaluators should be engaged. Such engagement of diverse groups helps in building a culture of evaluation in schools and aids in holding the system accountable. The entire system becomes accountable for itself when evaluation and development of the profession is everyone’s objective.
Significance of the Research

The study findings are significant on three distinct levels, teacher evaluation literature, teacher evaluation in Saudi Arabia, and for evaluation, measurement, and research literature. For teacher evaluation literature, this dissertation provides a detailed description of the Saudi teacher evaluation system. It also provides insight on the process of change that needs to be done through the alignment of practice with teaching standards. Studies about teacher evaluation in Saudi Arabia in general, and about alignment between the practice and the standards specifically, are scarce if not nonexistent. For teacher evaluation in Saudi Arabia, the findings of this study may aid in reform plans. Saudi teachers’ perspectives on teacher evaluation in the country have never been investigated and this input would likely be helpful when considering reforms. The deep thoughts and insightful experiences shared by teachers and some policymakers provide insight on a crucial issue which needs attention.

For evaluation, measurement, and research fields of inquiry, the study provides a contribution to methods and practice. The in-depth description of the teacher evaluation system and its examination through diverse lenses, should benefit policymakers responsible for teacher evaluation both inside and outside Saudi Arabia. Further, the framework for teacher evaluation articulated for this study could be adapted and utilized by others interested in more comprehensive models for studying teacher evaluation systems.

The manner in which the data was collected, analyzed, and presented informs and illustrates practices for implementing sequential mixed methods designs. Also, the development and refinement of the survey instrument used could aid other researchers in the field in measuring constructs about teacher evaluation, teacher working conditions, and teacher effectiveness.
Recommendations for Future Research

The focus of the current study was to investigate the alignment between teacher evaluation practice and the Saudi teaching professional standards. The limitations to the study have been discussed earlier in this chapter and in the methodology chapter. These limitations provide future research opportunities, some of which are presented here. The current study has investigated teacher evaluation in a public education system which stretches from 1st grade in elementary school to 12th grade in high school. It also looked at teacher evaluation in the whole country. Future researchers could investigate one level of education or a more specific part of the country to gain more insight on any issues that are related to these smaller populations. The study also tried to look at teacher evaluation through different lenses. Though it was successful, there still is a need to focus more deeply on other perspectives on teacher evaluation. For example, the phenomena of resistance to teacher evaluation could be investigated as a phenomenological study. Moreover, the perspective of key informants could also be better investigated in a specific study, if they are able to participate and discuss their visions. The perspective of other stakeholders, such as principals, superintendents, and parents could also be investigated for a broader coverage on the expectations, experiences, and needs of teacher evaluation in the country.

Besides looking at the practice of teacher evaluation, other studies could focus on the preparation and training evaluators receive before they participate in evaluating teachers. It is crucial that any teacher evaluation system has well-trained evaluators, as they are the individuals who will execute any evaluation plans. Having them well prepared and trained will raise the authenticity and validity of any gathered data.
One of the main struggles through this study was obtaining official documents about teacher evaluation in Saudi Arabia. A future study might try to obtain these documents and conduct a content analysis to investigate the underlying motives and purposes of teacher evaluation and its future in the country. It would also be beneficial to validate the current classroom observation tool that has been used for a long time without any validation. That would mainly depend on the availability of data that could be used for this purpose. However, since the Ministry has moved completely to an electronic system, data from teacher evaluations might be more readily available.

The survey that was used in this study needs more attention as well. A validation study focusing on the factorial structure and the reliability of the survey would benefit future researchers investigating related topics. Measurement specialists could also aid in building evaluation instruments that are capable of measuring teacher effectiveness based on the STS.

Lastly, there must be schools and communities of teachers who have been successful in implementing developmental teacher evaluation strategies somewhere around the country. Conducting case studies for such successful stories help in documenting and learning from these efforts. This could encourage other schools to try and adopt some of these strategies. The Ministry would also benefit from such studies in developing and improving the teacher evaluation system. Also, pilot studies implementing different interventions in regard to teacher evaluation at some schools that are willing to improve will help guide the future research and inform policy changes and future reforms.
Conclusion

In this chapter, the answers to the three research questions were briefly presented. Also, based on these answers, the findings, and the discussions in the previous chapter, a theoretical framework for teacher evaluation was presented. The framework links the inputs, procedure, utility, and outcome of the teacher evaluation system. The framework also aims to enhance the conceptual understanding of how the system works and building of the required tools to evaluate it. In addition to the framework, a set of policy options and recommendations that are based on the teacher evaluation literature and international best practices were presented.

The main aim of the dissertation was to investigate whether teacher evaluation practices are in alignment with the national teaching professional standards that were published in 2018 by the Education and Training Evaluation Commission (ETEC). The practice of teacher evaluation was required to be described and investigated first. Through the journey of doing so, a comparison between the practice of teacher evaluation and the international best practices was conducted. To sum up the findings, the practice of teacher evaluation in Saudi Arabia currently relies heavily on the subjective opinions of principals through an unvalidated classroom observation form. The form does cover a considerable number of the teaching professional standards; however, this addresses only “face validity.” There are also a few standards that are not covered in the observation form. Comparing the practices of teacher evaluation in Saudi Arabia with what is recommended by international best practices reveals that there are several methods and procedures that should be considered for the Saudi teacher evaluation system.

The recommendations in this dissertation were based on the international best practices and the teacher evaluation literature. The focus of the recommendations was on the importance of improving and institutionalizing teacher evaluations as one of many steps that can be taken to
improve overall outcomes of the Saudi education system. Teachers should always feel that evaluations are done for them and for the benefit of the teaching profession. When teachers feel they are targeted by evaluations to show their shortcomings they are more likely to assertively defy the evaluation system. It should be clear that the evaluation system does not prioritize being a threat to individuals in anyway. Thus, the developmental and formative aspect of the system should be prioritized and expanded. The attainment of such an objective will smooth the implementation of any developments and reforms in the education system, as teachers are more guaranteed to be onboard.
REFERENCES


Appendix A

The Current Teacher Evaluation Observation Form and Its Translation
### تقييم الأداء الوظيفي للمعلم

<table>
<thead>
<tr>
<th>الإدارة</th>
<th>المدرسة</th>
<th>المرحلة الدراسية</th>
<th>الهيئة</th>
<th>التخصص</th>
<th>المؤهل</th>
<th>التوجيهات والتوجيهات</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>الأداء الوظيفي</th>
<th>وصف</th>
<th>النهاية الظ Fridays</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>الالتزام باللائحة الفصحي</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>الحرص على تنظيم النشاط المدرسي وتقديمه</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>الاهتمام بالنمو المعرفي</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>المحافظة على وقت الدوام</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>الاهتمام باللائحة الفصحي في إعداد الدروس وتطبيقها</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>التمكن من المادة العلمية والقدرة على تحقيق أهدافها</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>الاهتمام باللائحة الفصحي ومراعاة الفروق الفردية</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>توزيع المنهج وتنفيذ ملائم ما تلقنه منه للزمن</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>استخدام المصادر والكتب المدرسية والوسائط التعليمية الأخرى</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>المهارة في عرض الدروس وإدارة الفصل</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>مستوى تصحيح الطلاب العلمي</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>الترمينات والإجراءات المدرسية والخليجية بتصحيحها</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>المجموع</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>الصفات الشخصية</th>
<th>وصف</th>
<th>النهاية الظ Fridays</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>السبب المتولد (النقطة المحددة)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>تقديم المسؤولة</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>تقديم التوجيهات</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>حسن التصرف</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>المجموع</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>العلاقات مع</th>
<th>وصف</th>
<th>النهاية الظ Fridays</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>الرؤساء</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>الزملاء</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>الطلاب وأولياء الأمور</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>المجموع</td>
</tr>
</tbody>
</table>

ملحوظات والتوجيهات

<table>
<thead>
<tr>
<th>وصف</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>قاد المدرسة</td>
<td>الختم</td>
</tr>
</tbody>
</table>
Teacher performance evaluation (Translated from Arabic)

<table>
<thead>
<tr>
<th>Administration</th>
<th>City</th>
<th>Level</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Specialty</td>
<td>Position name</td>
<td>Education level</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Job Performance points</th>
<th>Personal traits points</th>
<th>Communication points</th>
<th>Total points</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>16</td>
<td>12</td>
<td>100</td>
</tr>
</tbody>
</table>

**Job performance**

<table>
<thead>
<tr>
<th>Elements evaluated</th>
<th>Highest score</th>
<th>Teacher score</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using Classic Arabic</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention to extracurricular activities</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention to cognitive development</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence to work schedule</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiarity with approaches of lessons preparation and teaching</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery of the content</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention to continuous assessment and recognizing individual differences</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution of the curriculum content throughout the year</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The usage of the white board, textbooks, and other educational instruments</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skills in presenting lessons and managing the class</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students' achievement level</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giving homework and scoring them</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>72</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Personal traits**

<table>
<thead>
<tr>
<th>Elements evaluated</th>
<th>Highest score</th>
<th>Teacher score</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>General behavior (being a role model)</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appreciation of responsibility</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accepting all guidelines</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to make sound decisions</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>16</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Communication with**

<table>
<thead>
<tr>
<th>Elements evaluated</th>
<th>Highest score</th>
<th>Teacher score</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principals</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students and parents</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Comments and guides**

<table>
<thead>
<tr>
<th>Element of the evaluation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal signature</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

The Saudi Teaching Standards and Its Translation
<table>
<thead>
<tr>
<th>Domain</th>
<th>Primary standards</th>
<th>Secondary Standards</th>
</tr>
</thead>
</table>
| Professional ethics and responsibilities | Adherence to the Islamic tolerant ethics and professional responsibilities | Adherence to the Islamic ethics  
Strengthening patriotism and diverse cultures  
Meeting provisional ethics and educational polices and regulations |
|                                    | Continuous professional improvement                                               | Having a professional development plan and working on implementing it  
Improving professional performance according to the professional standards |
| Engaging with educators and the society | Engage with parents                                                                 | Engage with parents  
Engage with professional learning societies  
Engage with the local community |
| Professional knowledge             | Familiarity with the linguistic, quantitative, and digital skills                  | Comprehend audible and read text  
Ability to write clearly with right use of grammar  
Reading and speaking using correct language  
Basic knowledge about calculus and measurement methods  
Data collection, analyzing, and interpretation  
Information technology use and digital skills |
|                                    | Know learners and how they learn                                                  | Child development characteristics and its impact on learning  
Individual differences and its impact on learning  
How students learn  
Characteristics of students with special needs |
|                                    | Know the content and how to teach it                                              | Knowing the subject content  
Knowing specific instruction methods for the subject |
|                                    | Know the curriculum and the pedagogy                                              | The theory behind instruction methods  
Educational curriculums and it is evaluated  
General instruction methods  
Educational technology and resources |
| Professional practice              | Planning for teaching                                                             | Planning for instructional units and exercises  
Using a variety of instructional methods  
Using educational technology and resources  
Developing the shared dimensions in different curriculums  
Promoting critical and creative thinking skills |
|                                    | Create supportive and interactive teaching environment                           | Set high expectations from students  
Positively managing students’ behavior  
Preparing safe and attractive learning environments  
Efficient use of lesson time  
Establishing a communication culture which promotes learning |
| Assessment of students' performance | Building assessment instruments                                                   | Assessment application  
engaging students in assessment procedures  
utilization of assessment outcomes  
Assessment reporting |
Appendix C

The Teachers’ Survey (English)
Dear teacher,

The Education and Training Evaluation Commission encourages you to take part in this study.

My name is Suhayb Kattan. I am a PhD student at Western Michigan University. In my dissertation, I plan to compare teacher evaluation policies in Saudi Arabia to those of other countries. I will also build a framework for teacher evaluation in Saudi Arabia. These issues are of a great relevance to your daily work. Your valuable contribution in this survey will help in describing the current status of teacher evaluation in Saudi schools. All information you choose to share in this survey will remain anonymous. The survey will take less than ____ minutes to be completed. For more information you can use contact details listed below:

Suhayb Kattan

Email: s.kattan@etec.gov.sa

Phone: 0503223231

Thank you for your cooperation.

I agree to take part in the survey.

- Yes
- No

**Demographic information:**

1. Are you a:
   - Male
   - Female

2. How old are you? _______ Years

3. What is the primary subject you teach?
   - Mathematics
   - Science (Chemistry, Biology, Geology, Physics)
   - Social studies (geography, history)
   - Religion
   - Language
   - PA
   - English
   - Other: ________

4. For how many years have you been a teacher? ________ years

For the Arabic version of the survey you can email the researcher on sak.3231@gmail.com.

---

2 For the Arabic version of the survey you can email the researcher on sak.3231@gmail.com.
5. How many schools have you taught at? ________ schools
6. In which region do you currently teach? (Drop down with the 13 different regions)
7. Which school level do you teach?
   - Elementary school
   - Middle school
   - High school
8. Were you required to take the teachers’ test that is provided by the National Center for Assessment?
   - Yes
   - No
9. If yes, in which year did you take the test? ________
10. What was your score on the general test? ________

- **Teaching working conditions:**
11. How many lessons do you teach per week? _____Lesson
12. Roughly, how many hours per week do you spend preparing for lessons: _____ (Hours)
13. How many hours per week do you spend communicating with parents? _____ (Hours)

<table>
<thead>
<tr>
<th>Statement</th>
<th>Totally disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Class sizes are reasonable such that teachers have the time available to meet the needs of all students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Teachers have time available to collaborate with colleagues</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Teachers are allowed to focus on educating students with minimal interruption</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The non-instructional time provided for teachers in my school is sufficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Efforts are made to minimize the amount of routine paperwork teachers are required to do</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Teachers have sufficient instructional time to meet the needs of all students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
14. To what extent do you agree or disagree with the following statements: (Use of time)

- **Teacher evaluation:**
15. Are you aware of the teaching standards which were published by the ETEC in 2019?
   - Not at all aware
   - Slightly aware
   - Very aware
   - Extremely aware

16. How frequent are you officially evaluated at your school?
   - Once every couple year
   - Once a year
   - Twice a year
   - Other: __________

17. Do unofficial/informal teacher evaluations happen at your school?
   - Yes
   - No

18. To what extent do the following people contribute to your professional evaluation (Never, Once a year, 3-5 times a year, more than 5 times a year).

<table>
<thead>
<tr>
<th></th>
<th>Never a year</th>
<th>Once a year</th>
<th>2-5 times a year</th>
<th>More than 5 times a year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superintendent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
19. To what extent are the following methods of evaluation used in your school to evaluate teachers: (Never, to some extent, always)

<table>
<thead>
<tr>
<th>Method</th>
<th>Never</th>
<th>To some extent</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom observation</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Attendance</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Students grades and scores</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Lesson plans</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

20. After your formal evaluation is done, do you receive written feedback from your evaluator (principal/superintendent)?
   - Yes
   - No

21. How frequently do you get a chance to discuss your evaluation with your evaluator (principal/superintendent)?
   - Never
   - Rarely
   - Often
   - Always

22. In your opinion, to what extent does the evaluation procedures you went through reflect or cover the following:

<table>
<thead>
<tr>
<th></th>
<th>Not considered at all</th>
<th>Considered with low importance</th>
<th>Considered with moderate importance</th>
<th>Considered with high importance</th>
<th>I do not know</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Adherence to the Islamic ethics</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2. Promotion of national identity</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3. Promotion of cultural diversity</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4. Adherence to the professional</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>responsibilities as a teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Relations with the principal</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>6. Relations with other teachers</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>7. Relations with the community</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>8. The amount of professional development I participated in</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>9. Knowledge of my main subject</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>10. Knowledge of instructional practices</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>11. My linguistic skills in the classroom</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>12. Ability to meet student’s specific needs</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>13. Ability to prepare for lessons</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>14. Ability to create a communicative learning environment</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>15. The student’s assessment procedures you apply</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

23. In your opinion, to what extent were the following evaluation methods utilized in your most recent formal evaluation: (*I do not know, Not used at all, used with low importance, used with moderate importance, used with high importance*)
Concerning your formal evaluation, to what extent did it directly lead to the following:
(No change, small change, moderate change, large change)

<table>
<thead>
<tr>
<th></th>
<th>Not used at all</th>
<th>Used with low importance</th>
<th>Used with moderate importance</th>
<th>Used with high importance</th>
<th>I do not know</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student achievement or scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Retention or pass rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Students’ learning outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Student feedback on my teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Feedback from parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Observation of my classroom teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Innovative teaching practices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Relations with students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Classroom management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Teaching students with special needs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Student discipline and behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Having students from different backgrounds and cultures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Extra-curricular activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

24. Concerning your formal evaluation, to what extent did it directly lead to the following:
(No change, small change, moderate change, large change)
Concerning your formal evaluation, to what extent did it directly lead to the following:

*No change*, *small change*, *moderate change*, *large change*

<table>
<thead>
<tr>
<th></th>
<th>No change</th>
<th>Small change</th>
<th>Moderate change</th>
<th>Large change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Change in salary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. A financial bonus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Opportunities for professional development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Change in career advancement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Public recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Shift in your work position or responsibilities,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Role in school development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

25. Concerning your formal evaluation, to what extent did it directly lead to the following:

*No change*, *small change*, *moderate change*, *large change*

<table>
<thead>
<tr>
<th></th>
<th>No change</th>
<th>Small change</th>
<th>Moderate change</th>
<th>Large change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Your classroom management practices</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Your knowledge in your main subject field</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. A development or training plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Your teaching of students with special learning needs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Your handling of students' disciplines and behavior.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Your teaching of students from different backgrounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. The emphasis you place on improving your students’ test scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
26. How would describe the last formal evaluation you received:
   - It was judgmental about the quality of my work
   - It was suggestive of ways to improve my practice

   **Professional development:**
27. How frequent are you expected to take part in PD?
   - Never
   - 1-3 times a year
   - More than 3 times a year
28. How many days have you engaged in PD activities during the last year? ___ days
29. Have you ever done PD that you were not required to do?
   - Yes
   - No
30. If yes, did you have to pay for it?
   - Yes
   - No
31. What kinds of PD are available for teachers to pick from? *(Choose all applicable)*
   - In person courses
   - Online courses
   - Education conferences
   - Formal qualification programs
   - Observation visits to other schools
   - Peer observation
   - Participating in a professional development group
   - Reading and self-learning.
   - Other:
32. How are the areas of PD specified?
   - I choose
   - My evaluator chooses after doing the evaluation

   **Teachers as a professional community group:**
33. To what extent do teachers in your school engage in the following activities: *(never, once per year, 3-4 times a year, more than 4 times a year)*

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>1-3 times a year</th>
<th>More than 3 times a year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Attend staff meetings to discuss the vision and mission of the school.  
2. Take part in developing school curriculum  
3. Discuss with teachers of the same subject, the selection of instruction media.  
4. Exchange teaching materials with other teachers  
5. Teach jointly as a team in the same class  
6. Engage in professional discussion with other teachers about learning development of some students.  
7. Work with other teachers on students’ assessment to reach a common ground  
8. Attend team conferences for the age group I teach.  
9. Engage in professional visits to give recommendations and feedback to other teachers.  

- **Open ended questions:**  
34. From your point of view, what are the main challenges facing teacher evaluation in Saudi Arabia?  
35. Are there any suggestions that you would like to share which could improve teacher evaluation?
Appendix D

A Cross Walk Table to Align Items from Different Surveys
<table>
<thead>
<tr>
<th>Q.</th>
<th>Section/Construct</th>
<th>TALIS 2018</th>
<th>TALIS 2008</th>
<th>The study survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Background and Qualification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Gender</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Age</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Highest educational attainment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Qualification pathway</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Qualification vintage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Qualification elements and preparedness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Motivation to join the profession</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Career commitment to teaching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>3 Employment status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>4 Do you work at another school</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>5 - Number of schools worked at</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>6 employment status</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>7 Highest level of formal education</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>8 – year test takes</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>9 – Score on the test?</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Current work</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Employment status tenure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Employment status FTE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Teaching experience</td>
<td>9</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Work commitment in multiple schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Work commitment in multiple schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Special needs teaching status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Subjects taught</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Time distribution – total hours</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time distribution – teaching hours</td>
<td>8</td>
<td>10 (lessons instead of hours)/week</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------</td>
<td>---</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time distribution – non-teaching hours</td>
<td>8</td>
<td>11 hours/week</td>
<td></td>
</tr>
<tr>
<td>10 Years teaching at this school</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12 Communicating with parents (hours/week)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13 Working conditions (7 statements, agreement 4-point Likert scale)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 Awareness of new teacher standards</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 Number of times evaluated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 Unofficial evaluations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 People contributing to evaluation and frequency</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 methods of evaluation and frequency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19 Getting feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 Discussing the feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21 indicators covered by the evaluation (the standards)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22 Level of evaluation methods application/importance</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23 Evaluation consequences</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 Evaluation consequences</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 Type of evaluation</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26 Type of evaluation</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27 Evaluation consequences</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28 General issues about evaluation in school</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Professional development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26 frequency of PD</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>27 number of days of PD last year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>28 PD that are not required</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>29 Paying for PD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Time for PD during work hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>salary supplement for PD outside work hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Less formal PD activities and impact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Need of PD in different areas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Will to participate in more PD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>reasons not to participate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Participation in induction activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Formal induction provision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Involvement in mentoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Professional development types</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Professional development content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Professional development incentives / support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Professional development impact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Professional development impact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Professional development needs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Professional development barriers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>11 Methods participation and impact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>31 Who specifies the need</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Feedback types and sources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Feedback impact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Feedback impact</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Teaching in general**
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>32.</td>
<td>School’s team innovativeness</td>
<td></td>
</tr>
<tr>
<td>33.</td>
<td>Engagement in collaborative activities</td>
<td></td>
</tr>
<tr>
<td>34.</td>
<td>Self-efficacy</td>
<td></td>
</tr>
<tr>
<td><strong>Teaching in the target class</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>Target class students’ characteristics</td>
<td>34-43 Questions about teaching in a specific classroom.</td>
</tr>
<tr>
<td>36.</td>
<td>Target class special needs focus</td>
<td></td>
</tr>
<tr>
<td>37.</td>
<td>Target class subject focus</td>
<td></td>
</tr>
<tr>
<td>38.</td>
<td>Target class size</td>
<td></td>
</tr>
<tr>
<td>39.</td>
<td>Target class time distribution</td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>Satisfaction with classroom autonomy</td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>Target class disciplinary climate</td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>Core teaching practices in target class</td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Teaching practices – assessment</td>
<td></td>
</tr>
<tr>
<td><strong>Teaching in diverse environments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>Self-efficacy in multicultural classrooms (filter)</td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>Self-efficacy in multicultural environments</td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>Diversity practices (Filter)</td>
<td></td>
</tr>
<tr>
<td>47.</td>
<td>Diversity practices</td>
<td></td>
</tr>
<tr>
<td><strong>School climate and job satisfaction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.</td>
<td>School climate</td>
<td></td>
</tr>
<tr>
<td>49.</td>
<td>Student-teacher relations</td>
<td></td>
</tr>
<tr>
<td>50.</td>
<td>Job commitment / career plans</td>
<td></td>
</tr>
<tr>
<td>51.</td>
<td>Workplace wellbeing and stress</td>
<td></td>
</tr>
<tr>
<td>52.</td>
<td>Workload, student behavior, and complex teaching demands stress</td>
<td></td>
</tr>
<tr>
<td>53.</td>
<td>Satisfaction with the profession and this school</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Societal value of the profession</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>55.</td>
<td>Teachers spending priorities</td>
<td></td>
</tr>
</tbody>
</table>

**Teacher mobility**

<table>
<thead>
<tr>
<th></th>
<th>Academic mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.</td>
<td>Academic mobility – purposes</td>
</tr>
<tr>
<td>57.</td>
<td>Academic mobility duration of time abroad</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>32 Teachers community support</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>29 Personal beliefs about teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30 Practices of teaching frequency</td>
</tr>
<tr>
<td></td>
<td>31 teaching practices (personal and in school)</td>
</tr>
<tr>
<td></td>
<td>32 school management</td>
</tr>
<tr>
<td></td>
<td>33 Teaching practices (subject specific)</td>
</tr>
</tbody>
</table>
Appendix E

Teacher Interview Protocol
Teacher Interview Protocol

Teacher Evaluation in Saudi Arabia Relative to National and International Teacher Evaluation Standards and Best Practices

Introduction

Thank you for taking part in the survey and agreeing to participate in this interview. I am a Ph.D. candidate at Western Michigan University and this interview is part of my dissertation. I am interviewing you to better understand your experience with teacher evaluation in Saudi Arabia. I will be exploring in depth some of the initial findings of the survey you took part in.

- The interview is done remotely using the WebEx application.
- I will be taking notes, but also audio-recording the interview, so I do not miss anything.
- All information gathered will be transcribed and de-identified.
- The recording will be destroyed after it is transcribed and de-identified.
- I will be asking you 11 main questions.
- I expect the interview to take between 45-60 minutes.
- I am really interested in your experiences, so please answer with what you think, not what you think I want to hear!
- If at any time you feel uncomfortable, you may skip a question or ask to stop the interview completely.

Again, thank you for participating in this interview and sharing your experiences with teacher evaluation.

Teacher standards:

1- How familiar are teachers with the national teaching standards?
2- Are the national teaching standards something that teachers embrace and perceive to be important?

The evaluation procedures:

3- The survey results shows that school superintendents and principals as the main source of teacher evaluation feedback. To what extent you think they can reflect on what you are doing in the classroom? Could any other stakeholders be included in the evaluation process?
4- What do you think of the current evaluation form that is used to evaluate your performance?
5- What aspect of teaching do you think are not sufficiently covered in the current evaluation system?
6- How do you think evaluating teachers should take place?
7- What obstacles or challenges are in place that make teacher evaluation difficult?
   a. Time and frequency
   b. Training of evaluator (administrator) to do evaluation
   c. Clarity of expectations
8- Do you think teachers should be concerned about their evaluations and its consequences?

Professional development:

9- If an official tells you that evaluations are done to develop your professional skills. Do you agree or not? Why? Does the feedback lead to decisions regarding PD?

Closing questions:

10- Do you want to add any things to this interview?
11- Do you have any questions to me?

Thank you for allowing the time for this interview and sharing this valuable information with me.
Appendix F

Key Informant Interview Protocol
Key Informants Interview Protocol

Teacher Evaluation in Saudi Arabia Relative to National and International Teacher Evaluation Standards and Best Practices

Introduction

Thank you for agreeing to participate in this interview. I really appreciate that you took some time out of your busy daily schedule to meet with me. I am a Ph.D. candidate at Western Michigan University and this interview is part of my dissertation. I am interviewing you to better understand the policy of teacher evaluation in Saudi Arabia. I will be exploring in depth some of the initial findings I got from the teacher survey and interviews.

- The interview is done remotely using the WebEx application.
- I will be taking notes, but also audio-recording the interview, so I do not miss anything.
- You do not need to turn your camera on during the interview.
- All information gathered will be transcribed and de-identified.
- The recording will be destroyed after it is transcribed and de-identified.
- I will be asking you twelve main questions.
- I expect the interview to take between 30-45 minutes.
- If at any time you feel uncomfortable, you may skip a question or ask to stop the interview completely.

Again, thank you for participating in this interview and sharing your knowledge and experience about teacher evaluation.

Interview Questions

1. *The national teaching standards*, what role do you think they should have in the teacher evaluation process?
2. *About half the teachers on the survey indicated that they either never heard of the national teaching standards or that they are not familiar with them. Does this finding make sense? And how can you explain this finding?*
3. In general, how well is the current teacher evaluation system? Does it really do what it intends to do? *(Prompt by providing some examples from interviews that raises flags about the validity of the system)*
4. Are leaders (school principals) prepared in terms of their familiarity with the standards and their evaluation knowledge and capabilities, solely evaluate teachers in the current system?
5. **Professional development** is a key part of the national teaching standards which should be the foundation of the teaching evaluation system. How can teacher evaluation be connected to the training and development opportunities? *(Give some examples from teachers)*

6. How do you view the current system in terms of consequences and how evaluation findings are used? What needs to be changed or improved? *(The relocation part is fundamental, leaders giving high grades to teachers to get rid of them)*

7. *One purpose of teacher evaluation is to help teachers improve which is known as formative evaluation.* What is your view on formative purpose of evaluation in the current system? What needs to be changed or improved?

8. *Evaluation results are usually used to determine teacher effectiveness, promote teachers for bonuses, allow relocation of a teacher, evaluate schools or administrations.* What is your view on summative purposes of evaluation in the current system? What needs to be changed or improved?

9. *What are the main struggles and challenges the current evaluation system faces? (Underqualified school leaders, excessive power for leaders, students' performance in the evaluation of teachers, low and inconsistency of the frequency of teacher evaluations, and leaders being so busy).*

10. *Where do you ideally see the teacher evaluation system in five years from today? What are the steps needed to get to that? Are there any specific plans awaiting to be executed? (Choosing qualified evaluators (leaders), better incentives system, more frequent visits, school-based superintendents to maximize the formative part of the evaluation, self-evaluations, and distributing the evaluation power)*

11. Anything to add that is not mentioned during the interview?

12. Do you have any questions for me?

Thank you for allowing the time for this interview and sharing this valuable information.
Appendix G

Survey Reliability and Factor Analysis
Table 1
Indecies descriptive statistics

<table>
<thead>
<tr>
<th>No. of items</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working Conditions</td>
<td>642</td>
<td>7</td>
<td>1.00</td>
<td>4.00</td>
<td>2.18 0.61 0.82</td>
</tr>
<tr>
<td>Standards inclusion</td>
<td>634</td>
<td>15</td>
<td>1.00</td>
<td>4.00</td>
<td>3.36 0.61 0.93</td>
</tr>
<tr>
<td>Administrative change</td>
<td>642</td>
<td>7</td>
<td>1.00</td>
<td>4.00</td>
<td>1.64 0.79 0.88</td>
</tr>
<tr>
<td>Skills change</td>
<td>642</td>
<td>10</td>
<td>1.00</td>
<td>4.00</td>
<td>2.43 1.05 0.95</td>
</tr>
</tbody>
</table>

Table 2
Independent samples t-test comparing females and males

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working conditions</td>
<td>2.00 0.58</td>
<td>2.28 0.61</td>
<td>5.40**</td>
</tr>
<tr>
<td>Standards inclusion</td>
<td>3.53 0.51</td>
<td>3.28 0.64</td>
<td>5.00**</td>
</tr>
<tr>
<td>Administrative change</td>
<td>1.64 0.78</td>
<td>1.64 0.80</td>
<td>-0.05</td>
</tr>
<tr>
<td>Skills change</td>
<td>2.55 1.10</td>
<td>2.36 1.02</td>
<td>-2.13*</td>
</tr>
</tbody>
</table>

*Significant at p < 0.05, **Significant at p < 0.001

Table 3
Analysis of variance based on the subject taught for the four indices

<table>
<thead>
<tr>
<th></th>
<th>Within groups variation</th>
<th>Between groups variation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>df</td>
</tr>
<tr>
<td>Working conditions</td>
<td>233.74</td>
<td>633</td>
</tr>
<tr>
<td>Standards inclusion</td>
<td>233.00</td>
<td>625</td>
</tr>
<tr>
<td>Administrative change</td>
<td>395.06</td>
<td>633</td>
</tr>
<tr>
<td>Skills change</td>
<td>696.18</td>
<td>633</td>
</tr>
</tbody>
</table>

*Significant at p < 0.05, **Significant at p < 0.001
Table 4
Post hoc (Tukey HSD) comparisons between groups of subject teachers on working conditions

<table>
<thead>
<tr>
<th></th>
<th>Mean difference</th>
<th>SE</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math - Science</td>
<td>-0.27</td>
<td>0.08</td>
<td>0.034</td>
</tr>
<tr>
<td>Science – Arabic Language</td>
<td>0.33</td>
<td>0.08</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 5
Analysis of variance based on the school level for the four indices

<table>
<thead>
<tr>
<th></th>
<th>Within groups variation</th>
<th>Between groups variation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>df</td>
</tr>
<tr>
<td>Working conditions</td>
<td>239.41</td>
<td>639</td>
</tr>
<tr>
<td>Standards inclusion</td>
<td>235.52</td>
<td>633</td>
</tr>
<tr>
<td>Administrative change</td>
<td>404.55</td>
<td>641</td>
</tr>
<tr>
<td>Skills change</td>
<td>710.42</td>
<td>639</td>
</tr>
</tbody>
</table>

*Significant at p < 0.05, **Significant at p < 0.001

Table 6
Post hoc (Tukey HSD) comparisons between groups of school level on working conditions

<table>
<thead>
<tr>
<th></th>
<th>Mean difference</th>
<th>SE</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary – High</td>
<td>-1.50</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Middle – High</td>
<td>0.23</td>
<td>0.06</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 7
Factor loadings for the survey Likert type questions

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
<th>Factor 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.13 Ability to prepare for lessons</td>
<td>.867</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.12 Ability to meet student’s specific needs</td>
<td>.793</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.15 The student’s assessment procedures you apply</td>
<td>.775</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.14 Ability to create a communicative learning environment</td>
<td>.771</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.9 Knowledge of my main subject</td>
<td>.753</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
21.4 Adherence to the professional responsibilities as a teacher .726
21.6 Relations with other teachers .711
21.10 Knowledge of instructional practices .701
21.2 Promotion of national identity .613
21.1 Adherence to the Islamic ethics .613
21.7 Relations with the community .596
21.5 Relations with the principal .590
21.11 My linguistic skills in the classroom .577
21.8 The amount of professional development I participated in .472
21.3 Promotion of cultural diversity .423

24.8 Your student's assessment strategies .880 .349
24.5 Your handling of students disciplines and behavior. .867
24.10 Handling of students' problems and needs .861 .317
24.9 Use of different teaching methods .843 .373
24.2 Your knowledge in your main subject field .834
24.1 Your classroom management practices .808
24.3 A development or training plan .807
24.7 The emphasis you place on improving your students test scores .777
24.6 Your teaching of students from different backgrounds .751
24.4 Your teaching of students with special learning needs .602

32.3 Discuss with teachers of the same subject, the selection of instruction media. .814
32.4 Exchange teaching materials with other teachers .791
32.7 Work with other teachers on students' assessment to reach a common ground .777
32.5 Teach jointly as a team in the same class .720
32.6 Engage in professional discussion with other teachers about learning development of some students. .720
32.9 Engage in professional visits to give recommendations and feedback to other teachers. .689
32.2 Take part in developing teaching aids .688
32.8 Attend team conferences for the age group I teach. .304 .582
32.1 Attend staff meetings to discuss the vision and mission of the school. .312 .370

23.6 Shift in your work position or responsibilities, .761
| 23.4 Change in career advancement | .734 |
| 23.3 Opportunities for professional development | .720 |
| 23.2 A financial bonus | .719 |
| 23.7 Role in school development | .631 |
| 23.5 Public recognition | .618 |
| 23.1 Change in salary | .574 |

| 12.6 Teachers have sufficient instructional time to meet the needs of all students | .751 |
| 12.4 The non-instructional time provided for teachers in my school is sufficient | .677 |
| 12.2 Teachers have time available to collaborate with colleagues | .642 |
| 12.5 Efforts are made to minimize the amount of routine paperwork teachers are required to do | .642 |
| 12.1 Class sizes are reasonable such that teachers have the time available to meet the needs of all students. | .632 |
| 12.3 Teachers are allowed to focus on educating students with minimal interruption | .624 |
| 12.7 Teachers are protected from duties that interfere with their essential role of educating students | .546 |

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.
Loadings < 0.3 were suppressed
a. Rotation converged in 6 iterations.
Appendix H

WMU HSIRB Approval
Date: December 10, 2020

To: Gary Miron, Principal Investigator  
    Suhayb Kattan, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: IRB Project Number 20-12-05

This letter will serve as confirmation that your research project titled “Teacher Evaluation in Saudi Arabia Relative to National and International Teacher Evaluation Standards and Best Practices” has been approved under the expedited category of review by the Western Michigan University Institutional Review Board (IRB). The conditions and duration of this approval are specified in the policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may only be conducted exactly in the form it was approved. You must seek specific board approval for any changes to this project (e.g., add an investigator, increase number of subjects beyond the number stated in your application, etc.). Failure to obtain approval for changes will result in a protocol deviation.

In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the IRB for consultation.

The Board wishes you success in the pursuit of your research goals.

A status report is required on or prior to (no more than 30 days) December 9, 2021 and each year thereafter until closing of the study.

When this study closes, submit the required Final Report found at https://wmich.edu/research/forms.

Note: All research data must be kept in a secure location on the WMU campus for at least three (3) years after the study closes.