Teacher Evaluation in Kuwait - Evaluation of the Current System and Consideration of Risk-Based Analysis as a Principle for Further Development

ALMUTAIRI, TALAL,S,SH,S,A

Teacher Evaluation in Kuwait - Evaluation of the Current System and Consideration of Risk-Based Analysis as a Principle for Further Development

By

Talal S. Almutairi

Thesis submitted to Durham University in fulfillment of the requirements for the degree of Doctor of Philosophy

School of Education

Durham University

July 2016

Declaration

This thesis is as a result of my research and has not been submitted for any other degree in any other university.

Abstract

Evaluation is an essential component in education. Through evaluation, educationists can identify what has been achieved, what needs to be improved and what should be developed. Educational evaluation can be applied to several areas, such as curricula, students, teachers and schools in general. This study focuses on teacher evaluation and argues that teacher evaluation systems should be more effective in accurately determining teacher performance, should support the making of fair decisions in relation to sanctions or rewards and should support professional development. The aims of this study were firstly to analyse and evaluate the current teacher evaluation system in Kuwait, and secondly, to suggest an alternative teacher evaluation system based on a 'Risk-Based Analysis' approach. This study used a combination of quantitative and qualitative methods to gather data. In particular, a questionnaire was designed to collect teachers' perspectives on the current system and was distributed to 599 teachers in nine primary schools in three Kuwaiti educational districts. Interviews were conducted with nine head teachers and twelve inspectors in order to obtain more in-depth data regarding the current system. The study also carried out additional interviews with the same head teachers and inspectors, and held a focus group with 45 teachers, in order to probe their views concerning the proposed alternative system.

The results of this study have revealed that the actual purposes of the current system are primarily focused on achieving summative evaluation, while the desired purposes of teacher evaluation are to secure both summative and professional development. The current system most frequently uses observation to evaluate teacher performance; however, participants expressed a desire for a broader range of evaluation tools to be used. The study found that teachers more frequently have discussions with, and receive written feedback from, heads of departments as opposed to the other two evaluators (head teachers and inspectors). They rated the value of the discussion and written feedback from heads of departments as more valuable than that which is given by the other two groups of evaluators. The study also found that teachers' views were clustered, with some teachers indicating that the current system does not support them in their performance development and others indicating that it does. Finally, the study found that adopting the alternative system would improve the validity and reliability of teacher evaluation, would link teachers' performance with promotions and rewards while introducing sanctions for underperforming teachers, as well as would facilitate and promote professional development.

Acknowledgements

First, I would like to express my gratitude to almighty God, the Gracious and the Merciful, for giving me the ability and the willpower to complete this research.

I would also like to acknowledge all those people who have supported me throughout my PhD studies. Without their support and contributions, I would not have accomplished it. First, I would like to express my immense appreciation, sincere thanks, and profound gratitude to my doctoral supervisors, **Dr. Per Kind**, for his persistent guidance and excellent supervision. His consistently positive and encouraging words have always filled me with confidence to accomplish this research. His professionalism and expertise, as well as his frequent guidance and reminders, have kept me on the right track and enabled me to complete this thesis. **Professor Peter Tymms** too**,** for his direction, support, valued opinions and comments that have contributed significantly to the development of the thesis. I have been lucky indeed to be supervised by these two respected supervisors.

Sincere thanks and appreciation go to my father and mother, who have always encouraged me and advised me to do my best. My deepest gratitude to my sister **Muneera,** who is the first teacher in my life and, as a result of her teachings, I completed all educational levels, including this PhD thesis. Special thanks to my brother, **Fahad**, for his everlasting support which greatly influenced me to complete my studies. I offer my thanks too, to my brothers and sisters, for their encouragement, support and sustaining prayers for my success.

I also extend my acknowledgement to all those who took part in this study. Finally, I extend my appreciation to the School of Education at Durham University which offered me the chance to complete my PhD and benefit from its facilities.

Last but not least, I would like to thank all those who helped and advised me whom I have not been able to name.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

1.1 Introduction

Evaluation takes place in what people do in all aspects of their lives, whether at a scientific or social level. Through evaluation, people can determine their abilities and attempt to develop themselves. A good example is stated by Alaani, Maqdad, Aldousarie (2003) when a painter has finished drawing, he will stop and back away slightly so as to carefully consider the canvas; if there is no smile on his face we can assume that painter is unhappy and dissatisfied. The painter then goes back to the palate and colours to add some improvements. What has happened to make the painter feel dissatisfied? The operation carried out by the painter has, in fact, been an evaluation for judging the successful production of the painting. In the painters' mind, a lot of criteria will be taken into account during the drawing-board stage. These criteria may relate to several theories, including the theory of representation or the mixing of colours, with the eventual aim of offering work that expresses their aspirations. The painter's judgment has been reached through applying these criteria. They are a frame of reference with which to compare his/her canvas to determine how it approaches or distances itself from the desired goal. This work represents a genuinely practical evaluation.

In the educational field, evaluation is an essential component. It is through evaluation, e.g. the extent to which educational targets have been achieved, that strengths and weaknesses can be identified (Alnajar, 2010).

In this chapter, the educational context of Kuwait will be described briefly (for further details about the context, see Appendix 1). This chapter also includes the statement of the research, aims of the study, research questions, the significance of the study, the study rationale, and the outline of the thesis.

1.2 The Context of Kuwait

Nowadays, educating and teaching the Kuwaiti people is the full responsibility of the Kuwaiti government. According to the Kuwaiti Constitution in 1962, Article (40):

> Education is a right for Kuwaitis, guaranteed by the State in accordance with law and within the limits of public policy and morals. Education in its preliminary stages shall be compulsory and free in accordance with the law. The law shall lay down the necessary plan to eliminate illiteracy. The State shall devote particular care to the physical, moral and mental development of youth (Kuwait Constitution, 2008).

As a result of this constitutional provision, education has been freely extended to all levels and types of education for citizens, as is the case now in Kuwait, with the exception of private schools and private universities.

The Kuwaiti government controls education through the Ministry of Education [MOE]. The MOE designs national curricula, and organises educational services and facilities (public schools) so as to be the same throughout Kuwait. Moreover, the MOE accepts teachers to work in schools and determines teachers', head' and inspectors' roles (MOE, 2013a).

1.2.1 Education system

The school system in Kuwait is divided into four levels: kindergarten, primary, middle and high school. The schools are separate for girls and boys. The MOE divides public schools into six educational districts (Ahmadi, Asimah, Farwaniya, Jahra, Hawalli, Mubarak Al-Kabeer) according to the geographical distribution of Kuwait; each district is accountable to the MOE for its schools (MOE, 2013a).

According to the MOE (2013b) in 2004/2005, they changed the stages of schooling in Kuwait to 5 stages in primary school, 4 stages in middle school, and 3 stages for high school. Primary is for students between six and ten years old. It is compulsory for children to attend this level, otherwise there are sanctions for parents. For each schooling stage, students should study for one year and achieve 'pass' to move to the next stage. Students are taught basic skills and some specific subjects (National curricula): Arabic, English, Mathematics, Islamic studies, Science, Citizenship and a summary history of Kuwait (social studies), and Computer Science. There are also some subjects that are taught but on which the students are not examined, such as Art and Sports Education.

Each school has a department for each taught subject. In each department, there are approximately six teachers, more or fewer, depending on the needs of the department; for example, the number of classes in school. In each department there are one or two heads, and each school should also have a head teacher and two or three assistant heads of school. With regard to inspectors, they work in the Departments of Inspection in the MOE. Each subject that is taught in Kuwait has a main department of inspection. The main departments are divided into six sub-departments for each of the six districts.

1.2.2 The current teacher evaluation system

The civil service commission [CSC] has changed the employees' performance evaluation by Regulation number 36/2006. The MOE has changed the teacher evaluation system with regard to this decree, in 2006 and continues to the present day (Kuwait Teachers Society [KTS], 2010)

The purposes of teacher evaluation are: to determine individual teacher performance during the school year accurately and objectively, and to develop teachers' performance. Teacher evaluation is also used to make decisions about promotion (MOE, 2011; KTS, 2010), about increasing teacher's salary and annual bonus as noted in the Salary Scale of Teachers no. 48/2011 (KTS, 2012), or to make decisions about either dismissal or transfer to the administrative staff as sanction (KTS, 2010).

Within the Arab context, it is common for individual teachers to be evaluated every year by both internal and external evaluators. In the Kuwaiti system, every teacher is annually evaluated by three evaluators: the head of department, the inspector, and the head teacher. It is based on a written form of teacher evaluation by each evaluator. The final teacher evaluation is entered into the CSE's online portal by the head teacher, after agreement among three evaluators. Also, the head teacher has to print the final report of the individual teacher evaluation which is signed by the evaluators, and sent to the educational observers in districts in order to insert them in the teachers' record system in the MOE (MOE, 2011; KTS, 2010).

With regard to specific tools that are used to collect evidence about teacher performance during the school year e.g. classroom observation, student achievements etc., these are not mentioned in the policy of teacher evaluation. However, the teacher evaluation policy

stipulates that evaluators determine teacher's performance both inside and outside the classroom and use a standardised checklist for determining teacher's performance inside the classroom (KTS, 2010).

1.3 Statement of research

The contention of this study is that teacher evaluation should be effective, which in this sense means that teacher evaluation systems should effectively determine teachers' performance accurately, support making fair decisions in relation to sanctions or rewards, and support professional development.

From my knowledge and search, few studies have investigated the current teacher evaluation system in Kuwait. Alsanafi (2012) conducted research to evaluate the current system based on social science teachers' views in middle schools. The research found that professional development, and determining teacher performance were largely met by the system. However, teachers did not obtain monthly feedback from evaluators and the system was not appropriate in making decisions about sanctions and rewards. Other drawbacks were highlighted within this study: first, teacher evaluation was found to be too subjective, and second, the reports of individual teacher evaluations were kept confidential; in other words, the mid-year report and end-of-year evaluation report are not made available to the teachers themselves. Alsanafi concluded the study by recommending further studies on teacher evaluation and to address inspectors, head teachers, and different subjects at different levels of schooling.

Sabti (2010) also indicated some drawbacks of the current teacher evaluation system. First, there seems to be a lack of appropriate training and workshops to improve teachers' performance. Such training, he claimed, should be based on the teacher evaluation report, but this is rare. Second, teachers have no role in suggesting training courses and workshops. Third, the evaluation relies entirely on evidence collected through observation of teacher's performance by head teachers, inspectors, and heads of departments. In his study, the researcher recommended that teachers should attend training course or workshop every year or every two years and that evaluators should use a wider range of tools for teacher evaluation.

The current study has been built on the findings and advice from these two studies and aims to seek ways in which teachers can have a stronger in and derive more benefits from teacher evaluation. To achieve this aim, this researcher has listened to teachers, head teachers and inspectors to obtain their views about the current system and how to make the teacher evaluation system more effective. Furthermore, this researcher has introduced an alternative system for teacher evaluation based on a 'Risk-based analysis' approach, to obtain participants' reactions to it and ultimately find a new system that benefits teachers and takes their views into consideration. As teachers, head teachers, and inspectors possess practical knowledge of evaluating teachers, they know what areas of teacher evaluation need to be improved or developed, and they can assess whether the idea being developed is likely to work and be valid.

Finally, this study has focused on public schools and excluded private schools, since the MOE keeps track of private schools in line with the regulations for this type of education. But these schools are run by school owners or chief executive officers in terms of making decision about teachers, and some private schools apply different teacher evaluation systems that are unique to their schools. This researcher has also excluded parents and students from participating in this study. However, this research has addressed the extent to which teacher, head teacher, and inspector accept the involvement of parents and students in teacher evaluation.

1.4 The aims of the study

The aim of this study is firstly to analyse and evaluate the current teacher evaluation system in Kuwait using data from teachers, head teachers and inspectors. The objectives are:

a) To determine what purposes dominate the current system and to compare this to participants' desired purposes for teacher evaluation.
b) To identify the tools that are used in the current system, and to compare this to tools of teacher evaluation that participants think *should* be used.
c) To analyse the role of evaluators and how this is regarded by teachers in the current system.
d) To find out if and how the current teacher evaluation system is supporting the development of teachers' performance.

The second aim of this study is to suggest an alternative teacher evaluation system based on a 'Risk-based analysis' (RBA) approach to participants, and in order to probe their views about its potential for the improvement and development of teacher evaluation in Kuwait.

The reason for entering and evaluating teacher evaluation through teachers, head teacher and inspectors, is partly that they are familiar with the current system and how it operates. Teachers, for example, are in the best position to see what is most beneficial for them. Similarly, inspectors and head teachers are able to give valuable data about the system, as they are the evaluators. Contrasting views are expected, and only by finding a common ground is it possible to move forward to effective solutions.

1.5 Research questions

RQ1a: What contrasting views exist among teachers, head teachers, and inspectors regarding the intended and actual purposes of the current teacher evaluation system?

RQ1b: What purposes would the participants in the study prefer to dominate teacher evaluation?

RQ2a: What contrasting views exist among teachers, head teachers, and inspectors about the tools of the current teacher evaluation system?

RQ2b: What tools would the participants in the study prefer to see used?

RQ3: What contrasting views exist among teachers, head teachers and inspectors about the involvement of evaluators in the current teacher evaluation system?

RQ4a: To what extent do teachers regard the current teacher evaluation system as supportive in the development of teacher performance?

RQ4b: To what extent is the current evaluation system used to award promotions and rewards, and what are the teachers' views on this?

RQ5: What are the teachers', head teachers' and inspectors' views on the proposed alternative system based on a 'Risk-based analysis' approach?

In all these research questions, differences between gender, educational districts,

experience in teaching, and subjects will be considered as background variables.

1.6 Significance of the study

This study is significant for several reasons:

- This study will point out the desired purposes of teacher evaluation based on perspectives of stakeholders operating in the system.
- This study will help identify what tools of teacher evaluation, from users' point of view, should be used to make the teacher evaluation system effective.
- This study will bring out perspectives of stakeholders operating in the system about conflicts, and possible ways of overcoming these. In this way the research may have general value and contribute towards better use of teacher evaluation in Kuwait and elsewhere.

According to MOE (2013c), the MOE sets out a plan to develop education in Kuwait. One component of development plan for education in Kuwait is evaluation and measurement. Therefore:

- This study might help decision-makers in the Kuwaiti MOE by enabling them to hear from teachers, head teachers, inspectors about the current system and obtaining in-depth information about how it may be improved
- This study could be useful for drafting a method for improving teacher evaluation through drawing on an alternative system.

1.7 Study rationale

This study has several motivations. First, this researcher is motivated to make teacher evaluation better at identifying successful teachers and underperforming teachers, giving recognition to the teachers who deserve it, and helping all teachers to improve and to access more professional development.

Second, this study is intended to fulfil recommendations from previous studies in Kuwait through analysing the teacher evaluation system based on the perspectives of head teachers, inspectors, and teachers (in different level of school and taught subjects).

Finally, there is a personal motivation. This researcher has been sent abroad to study the field of educational evaluation and measurement by the MOE. Therefore, this researcher

set out to fill the gap related to teacher evaluation due to the lack of studies that evaluate teacher evaluation in Kuwait.

1.8 Outline of the thesis

Chapter one: this chapter is an introduction that includes the aims of the study, research questions, significance of the study and the rationale of this study. This chapter also includes a short brief overview of the context of Kuwait.

Chapter two: this chapter is a literature review that analyses teacher evaluation systems in general by looking at previous studies. Previous studies in the Kuwaiti context are also analysed.

Chapter three: this chapter presents the methodology and explains the study design, instruments that were used to collect data, the sample, and the quality of data.

Chapter four: in this chapter, the quantitative data about the current system collected from teachers through the questionnaires are analysed.

Chapter five: in this chapter, the qualitative data about the current system collected from head teachers and inspectors via the interviews are analysed.

Chapter six: this chapter describes the proposed alternative system that was introduced to participants in this study.

Chapter seven: this chapter explores participants' views on the proposed alternative system that was presented to them in focus groups and interviews.

Chapter eight: in this chapter, the results of this research are discussed and linked to previous studies; implications, recommendations, limitation, and further research conclude the thesis.

Chapter Two: Literature Review

2.1 Introduction

A review of literature can provide a rationale and foundation for a study, and also help further locate the focus of the inquiry. In this chapter, studies on teacher evaluation will be reviewed in relation to the main argument of this thesis, that teacher evaluation systems should be effective in accurately determining teachers' performance, should support the making of fair decisions in relation to sanctions or rewards, and support professional development. The consequences of an inadequate teacher evaluation system are twofold: there is little improvement in teachers' performance and there is ongoing employment of weak teachers (Donaldson & Peske, 2010).

From a review of a large amount of literature on teacher evaluation, the following appear to be key factors in an effective system:

- Maintaining a balance between professional development and summative purposes.
- Having explicit criteria for evaluating teachers' performance.
- Involving both external and internal evaluators.
- Using multiple tools to evaluate teachers' performance.
- Providing appropriate feedback: and
- Having qualified evaluators.

The chapter will first provide a conceptual framework for teacher evaluation, explain accountability, and analyse each of the listed factors from the research literature. Where possible, the chapter will bring in studies of teacher evaluation in Kuwait when reviewing these factors, but the chapter will also look at literature analysing the current and previous teacher evaluation systems in Kuwait in a separate section.

2.2 Conceptual framework of teacher evaluation

A teacher is one of several important elements present in schools integral to increasing the quality of education. In large part, this means ensuring that the teacher is highly skilled and able to perform to the best of his/her ability (Santiago & Benavides, 2009). In order to understand what comprises effective teacher evaluation, a more precise definition of the term is necessary. Teacher evaluation refers to the functions designed to make sound judgements about a teacher's performance and sound decisions about sanctions or rewards.

These functions should also encourage and assist teachers in developing their performance (Nolan & Hoover, 2008). According to Santiago and Benavides (2009), a range of interrelated questions are associated with teacher evaluation, such as who, with whom, what, how, and for what? (see Figure 2.1).

Who? Teacher evaluation is carried out on an individual teacher, but it is also a part of a broader evaluation and assessment that includes student assessment, school evaluation, and system evaluation. Kellaghan, Stuffelbeam and Wingate (2003, p. 1) explain the concept of this type of evaluation in a comprehensive manner:

> Educational evaluation encompasses a wide array of activities, including students assessment, measurement, testing, program evaluation, school personnel evaluation, school accreditation, and curriculum evaluation. It occurs at all level of education systems, from the individual students evaluation carried out by class-room teachers, to evaluations of schools and districts, to district-wide program evaluation, to national assessments, to cross-national comparisons of student achievement.

Teacher evaluation is therefore relative to school evaluation, and student assessment is relative to both school evaluation and teacher evaluation. According to Santiago and Benavides (2009), teacher evaluation can be interlinked with internal school evaluation since the results of teacher evaluation can be used to arrange for improvements to be made to the quality of teaching in different ways, such as part of an overall school-wide plan for improvement. Moreover, professional development activities for teachers, which are based on teacher evaluation results, can be linked to the general development plan for a school. Furthermore, an individual teacher's contribution to their school's development can be determined by their evaluation, since teacher evaluation covers a teacher's total contribution, such as their commitment to their professional development and their contribution to their school's management. Finally, students' results from a range of assessments can be used to make judgements about the school and its teachers. Assessment of learning is used to collect evidence about students' learning by determining whether they have understood what they have learnt, met program targets and achieved curriculum outcomes, and to make judgements about students in order to certify their proficiency (Earl & Katz, 2006). Here, assessment provides parents and students with information about their progress while also providing educators with information and administrators with

evidence to include in their reports on teachers or schools or another part of the educational system (Harlen, 2006). Assessment *for* learning, by contrast, is used as an investigational tool to discover learning (Earl & Katz, 2006) and to support learning (Gardner, 2006) through enhancing the feedback between teachers and students, adjusting teaching practice and pedagogical modes to meet students' needs and requirements, and to motivate students in their learning and improve their self-esteem (Black & Wiliam, 2006).

With whom? This aspect concerns the implementation of teacher evaluation; it relates to the involvement of a range of stakeholders in developing and conducting teacher evaluation, which can include students, teachers, parents, head teachers and administrators (Santiago & Benavides, 2009). The evaluators who are involved in conducting teacher evaluation can be divided into internal and external evaluators. Internal evaluators might include a teacher or group of teachers, members of the school, such as professional staff, head teachers, and/or other administrative members (Nevo, 2001; Ryan, Chandler, & Samuels, 2007). Internal evaluators might also include students (MacBeath & McGlynn, 2002). Examples are given below:

- *Principal/Head Teacher or Head of Department:* Principals, head teachers, or heads of departments may be asked to use their full contextual knowledge about their schools, students, and teachers/teaching to evaluate their teachers (Goe, Bell, & Little, 2008).
- *Teacher:* A teacher can participate by encouraging another teacher to evaluate his/her personal teaching (Airasian & Gullickson, 1997) or by evaluating each other, or by evaluating himself/herself (Nolan & Hoover, 2008; Arnodah, 2013).
- *Student:* Students can be allowed to rate their experiences with their teachers based on their interactions (Goe et al., 2008).

An external evaluator is usually a person who does not work inside the school, such as an inspector or consultant (Vanhoof & Van Petegem, 2007). In many countries, inspectors or professional evaluators working for national or regional educational authorities conduct external evaluations. Evaluation agencies working on behalf of the school or government can also conduct external evaluations (Nevo, 2001). Furthermore, teachers can be external evaluators when they participate in the evaluation of teachers working in other schools (Santiago & Benavides, 2009).

What? As described by Santiago and Benavides (2009), this aspect concerns the scope of teacher evaluation, covering areas such as: planning and preparation (e.g. knowledge of content, pedagogies, selecting educational targets, demonstrating student knowledge, assessing student learning, etc.); the classroom environment (e.g. creating an environment for learning, managing the class, dealing with student behaviour, etc.); instruction (e.g. using clear and accurate language, questioning techniques, discussion techniques, the students' engagement in learning, providing feedback to students, etc.); and professional responsibilities (e.g. demonstrating professionalism, professional growth, communicating with families, contributing to the school, etc.).

How? This aspect concerns the tools of teacher evaluation used to identify a teacher's performance in relation to certain criteria (Santiago & Benavides, 2009). More precisely, criteria for teacher evaluation constitute benchmarks for evaluating the performance of an individual teacher (Nolan & Hoover, 2008). In other words, they set "standards to evaluate teachers relatively to what is considered as good teaching" (Isore, 2009, p. 11). The tools of teacher evaluation include classroom observation, self-evaluation, student evaluation, peer evaluation, teacher portfolio, and student achievement. *Classroom observation* is a technique used for collecting data about teachers, students, and the relationship between them with respect to learning (Goe, Biggers, & Croft, 2012). Evaluators normally use a checklist during observation and look at two elements: the teacher's performance in class and the students' understanding and participation during particular tasks (Montgomery, 1999). *Self-evaluation* means asking teachers to write reports about themselves that include information about aspects of teaching, such as their teaching techniques, subject areas, etc., and, more importantly, how they are doing it and why (Little, Goe, & Bell, 2009). Self-evaluation can also be used in teacher evaluation during professional conversations/interviews or surveys/teaching logs in order for evaluators to discuss the information and details provided by teachers with them in a productive manner (Mather, Oliva, & Laine, 2008; Little et al., 2009). *Student evaluation* is used to collect data about teaching practices (Little et al., 2009). Surveys have the potential to gather the students' input at different levels, that is, primary, middle, and high schools (Ferguson, 2010). Student evaluation can also be conducted by focus group interviews to ascertain their opinions and feelings about their teachers' teaching practices (Nolan & Hoover, 2008). In *peer evaluation*, teachers are required to evaluate their colleagues (Arnodah, 2013) through peer observation (Nolan & Hoover, 2008), or by examining documents, such as lesson

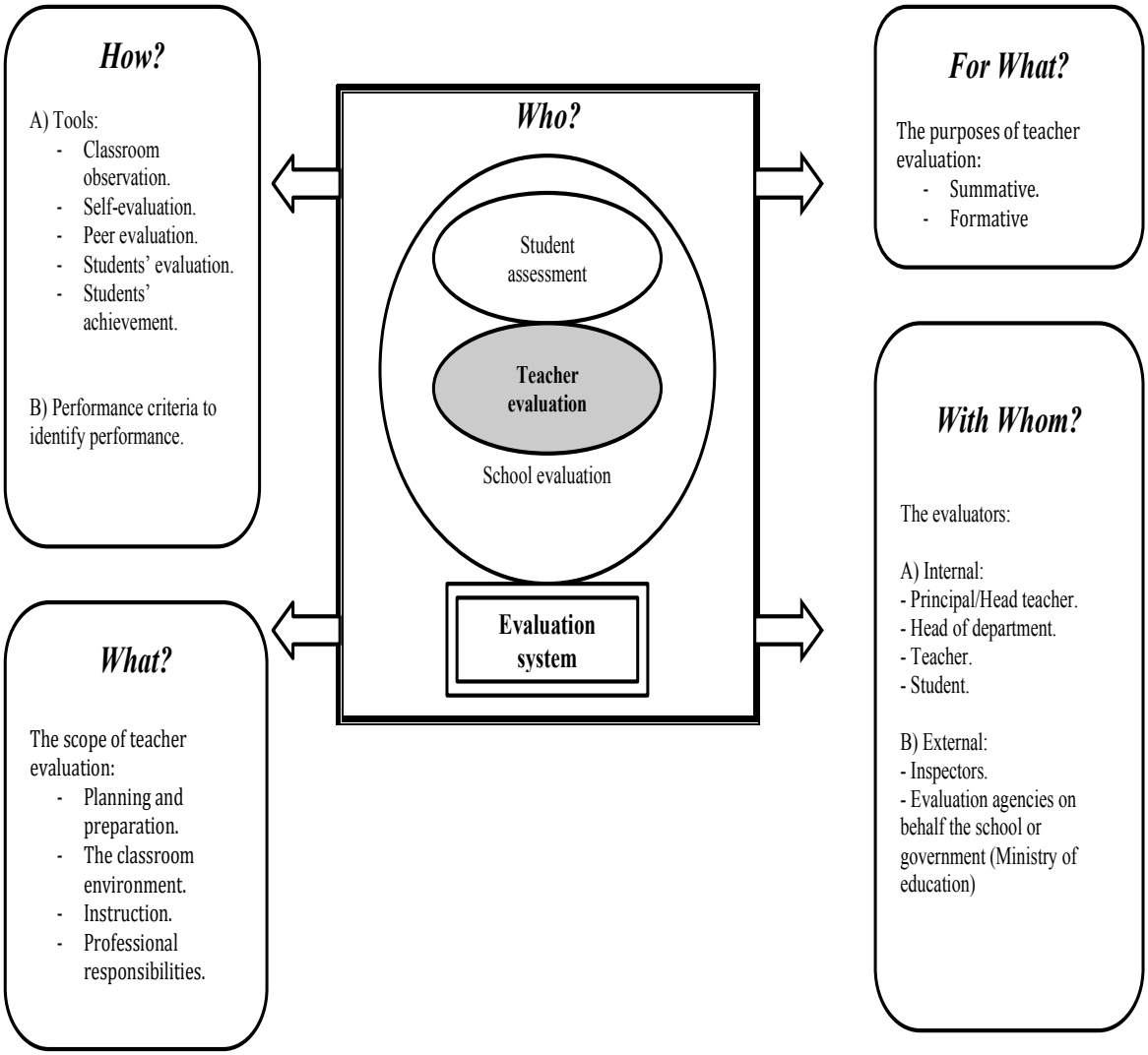
plans, assignments, and other evidence about extracurricular activities (Joshua, Joshua, Bassey, & Akubuiro, 2006). *Teacher portfolio* is a tool for evaluating teachers that includes a range of materials that are collected by teachers to be used as evidence of their teaching practice, their contribution to school activities, and their students' progress. Portfolios might also include such items as a sample of their students' work, assessments, lesson plans, and schedules (Little et al., 2009). *Student achievement* data from, such as standardised tests can also be used to determine the teacher's contribution to student learning, and such systems use statistical models to analyse the change in standardised test results over time (e.g., by comparing the current year to the previous year) (Mather et al., 2008).

For what? This aspect concerns the purposes of teacher evaluation with regard to the mechanisms for using the results. There are two major ways to classify the purposes of teacher evaluation; they can be either professional development or summative purposes (Santiago & Benavides, 2009; Stronge, 2006).

In particular, summative teacher evaluation is used to determine the "merit, worth, or value" of a teacher's performance (Smith, 2001, p. 51). Determining this performance may be based on a teacher's level of knowledge, communication of content, concepts of education, teaching and monitoring capabilities, occupational morality, extent of assisting colleagues and collaborating with them, teacher-student interaction (Cai & Lin, 2006), and their ability to achieve the required educational targets (Isore, 2009). Summative evaluation is also used to decide on sanctions and rewards. It can determine career advancement/promotion, award performance rewards, such as higher scale or bonus pay, or establish sanctions for underperforming teachers, leading to dismissal or delaying career progression (Davidson, Jensen, Klieme, Vieluf, & Baker, 2009).

Formative teacher evaluation identifies the different ways that teachers can develop their practice (Mathers et al., 2008) and therefore has professional development as its focus (Stronge, 2006). Teachers can identify their areas of strength and weakness, and learn strategies for how to improve and enhance their professional development (Stronge, 2006; Santiago & Benavides, 2009).

Figure 2.1: The conceptual framework of teacher evaluation



- Santiago & Benavides (2009).

2.3 Accountability

An accountability system is used to focus on student achievement or test results to monitor school performance (Gurr, 2007). In other words, accountability system is conducted to provide information to decision-makers and the public about learning outcomes to ensure that educational targets are being met with effective use of recourses (Faubert, 2009). An accountability system may also generate sanctions and rewards for schools based on student outcomes (Figlio & Loeb, 2011). In this sense, schools are held accountable to several authorities such as ministries, councils, stakeholders, and parents. Conceptually, there are three dimensions of accountability that are related to school evaluation: contractual, moral, and professional accountability as described by Gurr (2007) and

Faubert (2009). *Contractual accountability* is the extent to which a school meets requirements, serves the community in terms of student learning, and in some contexts publishes league tables and school guides. *Moral accountability* is concerned with meeting the needs of parents and students and ensuring a safe and high-quality school environment, as well as in some contexts offering parents the opportunity to choose a school based on the quality of education. *Professional accountability* is concerned with meeting the school's own expectations and those of other schools in order to add significantly to the learning of students to become the best schools in the district or the state—i.e., leading to change and improving student outcomes. To illustrate this point, using evidence to support schools in concentrating their attention on curriculum and instruction to improve and develop student learning (Figlio & Loeb, 2011).

The process of evaluating schools based on student performance via standardised tests is increasingly prevalent around the world. Many countries apply the accountability system such as the USA, United Kingdom, the Netherlands, and Australia (Figlio & Loeb, 2011; Faubert, 2009; Gurr, 2007). English schools, for example are accountable to local authorities and stakeholders such as parents through OFSTED inspection. Local authorities and the Secretary of State have a duty to intervene if school performance becomes a cause for concern. Parents also have an opportunity to choose the school for their children based on the information provided to them about schools' rankings or results. In the Netherlands, schools are accountable to central government for both budget matters and student achievement. The ministry in the Netherlands also has the right to restrict schools. School are also accountable to parents, as they have the freedom to choose a school based on inspection reports and students achievement (Faubert, 2009). The USA has the most famous federal system—that is, the No Child Left Behind Act [NCLB]. This system requires states to test students in reading, mathematics, and science, evaluating schools based on their students' outcomes (Figlio & Loeb, 2011).

Many previous studies have investigated the effect of an accountability system on student achievement. Wong, Cook and Steiner (2009) conducted a study to evaluate NCLB in the USA using National Assessment of Education Progress data between 1990 to 2009 for math in grades 4 and 8. They found positive effects of the accountability system on student achievement for both grades. Cronin, Kingsbury, McCall, and Bowe (2005) analysed longitudinal data on student achievement before NCLB (2001-2002) and after

implementation (2003-2004). They found that math and reading exam scores improved over the two years since NCLB was implemented. Neal and Schanzenbach (2010) found that after NCLB, students' scores in reading and math increased among students in the middle of achievement distribution, but not among students in the least academically achieving Chicago Public schools. They suggested that teachers tended to pay more attention to students who are near the proficiency standard. Ladd (1999) conducted a study to measure the effect of accountability on student outcomes after Dallas implemented an accountability system; in this study, the researcher compared Dallas student outcomes to outcomes of students in other districts. The researcher found that the passing scores in Dallas increased after implementing accountability, compared to other Texas districts. While, Smith and Mickelson (2000) compared outcomes from three North Carolina districts, of which one of three (Charlotte-Mecklenburg) had implemented an accountability system, and those researchers found there was no evidence of any effect on achievement. In terms of schools, Rockoff and Turner (2008) found that after conducting accountability, failing schools saw positive effects from accountability pressures in New York, as those school improved to the "D" level. Altrichter and Kemethofer (2015) conducted research in seven European countries (the Netherlands, England, Sweden, Ireland, the Czech Republic, Austria, and Switzerland) using online survey data from 2300 principals. They found that principals who feel more accountability pressure pay closer attention to the quality expectations of inspections, are more sensitive to stakeholders' reactions to school results, and are more engaged in improvement. With regard to teachers' views on testing students and using the results in an accountability system, Hamilton, Berends and Stecher (2005) collected math and science teachers' responses from primary and middle schools math in three states (California, Georgia and Pennsylvania). They found that teachers are engaging in a number of professional development activities to align their teaching with state standards and tests, and the accountability system has an effect on schools and teachers of focusing on student learning to meet the targets.

In the context of Kuwaiti, from this researcher's experience and search in Education Act, there is no accountability system at this point. Public schools are not evaluated based on student achievement, and there is no school rank list, but schools are run and controlled by the MOE (see Appendix 1). Furthermore, parents are not allowed to choose schools for their children based on the quality of school; instead their children are registered with the school that is nearest to their home address.

2.4 Maintaining a balance between professional development and summative purposes

The first factor to be looked at is the need for teacher evaluation to strike a balance between professional development and summative purposes.

Delvaux et al. (2013) conducted a study on the teacher evaluation system in the Flanders. They used a questionnaire, and their study's sample included 1983 teachers in 65 schools. An important outcome was that although the intended purpose of the teacher evaluation system was formative, i.e. for professional development rather than summative, the actual findings showed no significance of this system with respect to the effect on teachers' development. The summative purposes of the system, in contrast, had a small but significant positive effect on teachers' development. The reason may be that teachers feel under pressure where evaluation is summative and may feel compelled to undertake development. Based on this study, it may *seem* as if effective teacher evaluation should focus on summative evaluation of teacher performance.

Peterson and Comeaux (1990), who conducted a study on teacher evaluation systems in Florida and Wisconsin districts using 48 teacher interviews and questionnaires, found that teachers gave a high rating to an alternative system which encouraged teachers to reflect on their own teaching, and concluded that their perspectives were influenced by the format of the alternative system. Teachers in their study saw the system as reflecting their teaching and its purpose as being mainly for professional development; these teachers believed the ideal purpose of teacher evaluation was to promote professional development. Peterson and Comeaux suggested that teacher evaluation systems serve several needs and that summative evaluation should be used to identify whether performance deserves sanction or reward while the professional development is to meet the needs for teacher improvement and development. Colby, Bradshaw, and Joyner (2002) also reviewed the research literature and found that effective teacher evaluation systems have two purposes: summative and professional development. Stronge (2006) also argued that teacher evaluation should commit to both professional development and summative purposes to productively serve the needs of individual teachers and the school as a whole.

Based on a study that included 15,401 teachers, 932 principals, and 831 other evaluators via online questionnaires in Tennessee, the State Collaborative on Reforming Education

[SCORE] report (2012) also supports the notion that an effective system of teacher evaluation maintains a good balance between both purposes. The aim of the study was to obtain feedback about a new teacher evaluation system in the state in comparison to an old teacher evaluation system. The old system offered no meaningful feedback to teachers, which was the reason for introducing the new system. In addition, while in the old system there was no requirement to make personnel decisions such as tenure or dismissal, the new teacher evaluation aimed to serve both formative and summative purposes. All teachers received annual evaluation in order to provide them timely feedback about their teaching, and in order to inform decisions about assignment, reward, promotion, and compensation. The new system was found to be supportive of effective teaching, encouraging self-reflection, and motivating collaboration among teachers. Moreover, principals and evaluators feel the new system is having a positive impact on student achievement. Principals and other evaluators frequently indicated in the roundtable sessions that the framework of the new evaluation system facilitates them in carrying out their work more effectively and being instructional leaders. Therefore, principals and evaluators agreed in questionnaires that the system would have a positive impact on instruction and student achievement in their school.

According to Davidson et al (2009) report on the Organisation for Economic Co-operation and Development statistics for 23 countries that participated in the first round of TALIS, the findings showed that teachers who received judgement and feedback, evaluation had a positive impact not only on their career, but also on their teaching. The greatest impacts were on student scores, classroom management, understanding of teaching practice, and knowledge and development or training plans.

2.5 Clarifying the criteria for teacher evaluation

A point repeatedly made in the research literature is that in order to make teacher evaluation better informed, criteria need to be made explicit (Nolan & Hoover, 2008), because teacher evaluation criteria explain what teachers should do (Philips & Weingarten, 2013). The absence of explicit criteria, Nolan and Hoover (2008) argue, will lead to personal whims guiding evaluators.

Alhamdan (1998) investigated Kuwait's teacher evaluation system through questionnaires given to teachers, inspectors, and head teachers. He found that some criteria used to

evaluate teachers were not made explicit or were in need of clarification. These criteria led evaluators and teachers to have different interpretations, and thus were difficult to measure. Alsanafi (2012) conducted a study to evaluate the teacher evaluation system in Kuwait by using questionnaires with teachers. The results of this study concurred with those of Alhamdan's research: unclear criteria can lead to different interpretations by evaluators and teachers, making it difficult to evaluate exactly what is expected from teachers' performance. Accordingly, if the criteria are not explicit, it will affect whether teacher evaluation can determine teacher performance accurately, which will in turn affect teachers' self-conception and understanding of what they need to do to develop their performance.

Similarly, the SCORE report (2012), in the US found that a teacher evaluation system that uses explicit criteria leads to a better understanding of effective teaching, and hence inspires better performance. A teacher evaluation system with clear criteria will outline to teachers what is expected of them and what they need to demonstrate. Zhang (2008) conducted a study in three schools in Shanghai, to examine the implementation of teacher evaluation in these schools. The researcher used interview, participant observation, and documents analysis. The sample consisted of 74 participants, including school leaders, middle managers, heads of departments, and teachers. The researcher found that explicit criteria provided directions for teachers to follow to meet objectives, and to compare their practice with these criteria. Explicit criteria lead to accurate evaluation, and can facilitate teachers' development. Delvaux et al. (2013) also found that teacher evaluation that includes explicit criteria has a greater impact on teachers' development, as explicit criteria motivate teachers to bring their performance into line with expectations.

Furthermore, a shared understanding between evaluator and evaluatee of criteria against which the performance is judged is necessary. The continued emphasis on explicit criteria is not sufficient to yield a shared understanding between evaluatee and evaluators (Rust, Price, & O'Donovan, 2003). If both evaluators and evaluatee have a shared understanding of criteria, evaluators become accustomed to using terms and can easily presume that the evaluatee knows what the evaluators mean, using appropriate terms to explain judgement of quality, and evaluatee can thus recognise the low or high level of their work or performance (Sadler, 2010). However, if they do not share their conceptions of the criteria, then the information that the evaluatee receives is unlikely to be useful (Hounsell, 1997).

In order to build a shared understanding of criteria between evaluators and evaluatee, the criteria should become a part of each evaluation vocabulary to enable them to rehearse in their mind as they arrive at a judgement, and later explain and justify that judgement (Sadler, 2010). Providing better definitions of criteria and performance-level definitions, increasing discussion and reflection between them about the criteria, and collaborating to devise and negotiate their own criteria for performance (Nicol & Macfarlan-Dick, 2006).

2.6 Involving internal and external evaluators

Another argument made is that effective evaluation systems need internal and external teacher evaluation.

Using internal evaluation has the advantage of encouraging schools to assume their own responsibilities or duties (Nevo, 2001). This is where internal evaluation comes under the freedom of schools, allowing them to take responsibility for evaluation and to come up with their own improvement plan (Vanhoof & Van Petegem, 2007). Furthermore, as Nevo (2001, p. 97) has indicated, "developing an internal evaluation mechanism in a school is also an investment in an enduring resource for serving the information needs of the school by means of data pools and school portfolios". Moreover, while teachers may feel under pressure and stressed during an external evaluation (Faubert, 2009), internal evaluation tends to be less threatening and might therefore reduce stress. Internal evaluators can help reduce teachers' feelings of being threatened, since they know the local problems, communicate better with those being evaluated, and are present at school to facilitate the implementation of the evaluation recommendations (Nevo, 2001). Finally, teachers could be encouraged to be involved in the decision-making and this can foster collaboration among teachers in reflective practices, which develop teacher professionalisation (Nevo, 2001). However, by conducting only internal evaluation, schools may hide problems from parents and external stakeholders. Furthermore, schools may set their own standards for quality that may not reflect high quality and the bar may be set too low.

The advantage of external evaluation is that it is conducted by someone who is not involved in the school; when an evaluation is conducted from outside the school, it can be viewed as being more valid where internal evaluation might be suspected as biased and subjective (Nevo, 2001). External evaluation can be biased too, but an external evaluation is less likely to be seen as subjective than an internal evaluation. Furthermore, while external evaluation is about stimulating commitment, internal evaluation can be conducted

by means of the external evaluation's requirements for determining whether schools are fulfilling their duties (Nevo, 2001). To illustrate this point, when external evaluators require information from a school regarding its internal evaluation of the quality of teaching or individual teachers' performance, they will do so by stimulating schools to be committed to conducting an internal evaluation that has and uses that information. Also, external evaluation often expands the scope of evaluation while internal evaluation might suffer from a narrower perspective on overall qualities; external evaluation "can add commonalities to the uniqueness of the school and also provide a basis to judge its qualities", as well as expand the scope, such as by comparing performance across schools (Nevo, 2001, p. 98). On the other hand, external evaluation is often focused on commonalities and comparability, so might be insensitive to issues particular to a school or teacher. Internal evaluation can assist the external with in-depth information to reflect the character of a particular performance, and add local perspective to the findings of the evaluation (Nevo, 2001).

Accordingly, using both internal and external evaluations is necessary for teacher evaluation to be effective. Vanhoof and Van Petegem (2007, p.108) argued that "if one is absent, the other loses value", echoed by Nevo (2001, p. 101) who stated that each can learn something from the other. He also makes the point that any "evaluation (internal and external) has to be modest, acknowledging its limitations".

2.7 Using multiple tools to evaluate teacher performance

Using multiple tools to collect data to evaluate teacher performance is another way to make a system effective (Colby et al., 2002; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012a; Stronge, 2006; Kane & Staiger, 2012).

There are several benefits to using multiple tools for teacher evaluation. Perhaps the greatest benefit is that it gives evaluators a better picture of the multifaceted elements of teaching practice, since they can take into consideration the full range of a teacher's performance throughout the school year (Burnett, Cushing, & Bivona, 2012; Lachlan-Haché, 2011). For example, Kane and Staiger (2012) found that using observation, student evaluation, and value-added data created a statistically stronger determination of effective teaching than observation alone. Moreover, Burnett et al. (2012) indicated that using multiple tools can also facilitate and support teachers to identify their performance strengths and weaknesses. The use of multiple tools can increase the amount of feedback

that a teacher receives and improve its quality since teachers are able to obtain specific feedback on their teaching in a different way in order to help them develop professionally (Burnett et al. 2012). Furthermore, using multiple tools allows evaluators, school and districts to make fair decisions regarding rewards or sanctions and improves the ability of evaluators to make decisions (ibid). According to Kane and Staiger (2012), by using multiple tools, decision makers will obtain better information than they would with one tool; thus, more complete information can facilitate making better decisions about such things as promotions.

The benefits of multiple tools in teacher evaluation can therefore be summarised as allowing evaluators to: better and more accurately determine teaching practice; make fair decisions; and provide higher quality feedback to improve and develop a teacher's performance. These benefits emanate from the information that is collected via multiple evaluation tools. As DePascale (2012) pointed out, each tool has a margin of error that may affect the reliability of the data; therefore, using multiple tools can increase their reliability. Kane and Staiger (2012) found that combining multiple tools (observations, student evaluation, value-added data) led to greater reliability as they found more stability in the data collected using those tools. Similarly, Hanover Research (2012) indicated that multiple tools lead to increased reliability (consistency) because evidence from various tools, which includes input from different perspectives, increases the possibility of corroboration with the other tools. Furthermore, Zhang (2008) found that data from multiple tools can validate one another and reduce bias resulting from a single tool, as observed in three schools in Shanghai that conducted evaluation using multiple tools. Hanover Research (2012) also confirmed that multiple tools increase the validity of the evaluation because they increase the number of performance components that are evaluated, and thus offer more accurate information about performance than a single tool.

To gain most benefits from using multiple tools, the selection of the tools for evaluating teachers should be based on the purposes of the teacher evaluation and the data's intended use (Leo & Lachlan-Haché 2012). For example, if the purpose of teacher evaluation is formative, then schools or districts should select the tools that provide specific feedback, whereas if the purpose is summative, then schools or districts should select tools that collect accurate and consistent data. Furthermore, there are different approaches for combining multiple tools that should be considered. A *numerical approach* uses various

tools and each tool is given a weight which is either an equal weight (50/50) or weighted based on the performance element it intends to evaluate (e.g., observation 50%, student achievement data, 15%, etc.) in order to generate a final teacher performance score (Leo & Lachlan-Haché, 2012; Harris, 2013; Hansen, Lemke, & Sorensen, 2014). According to Harris (2013), this approach is commonly used for summative purposes. One advantage of this approach is that a teacher can avoid being penalised for a weak score in one area of evaluation by compensating with a higher score in an area of strength (Leo and Lachlan-Haché 2012). For example, a teacher can be evaluated as outstanding in their overall score according to observation and student achievement even though the teacher is rated weak based on student evaluation. This approach can also reduce the effect of single-tool bias (Leo and Lachlan-Haché 2012). For example, some head teachers give high scores for classroom observation to avoid having low school performance scores, so the results found by using other tools can alter the biased data in the overall score. Leo and Lachlan-Haché also argue that this approach is helpful to determine the weight and composition of groups of teachers (e.g., new teachers compared to more experienced teacher) since this approach means the system can be flexible. For example, for new teachers, the system can give classroom observation a greater weight than the other tools. However, this approach may lead to misclassified results, since low performance scores may be classified as good or very good according to other weights, as indicated by Harris (2013). Also, nuances can be lost by using a single score, and thus teachers may not get the feedback that is needed for their improvement (Leo and Lachlan-Haché, 2012).

Another approach is to use a *portfolio* or *matrix*; in this approach, teachers are evaluated using multiple tools, and each tool is considered and scored separately before the data is combined to determine their overall scores (Leo & Lachlan-Haché, 2012; Harris, 2013; Hansen et al., 2014). For instance, evaluators rate teachers on each performance aspect and provide their different views on teacher performance; then, using the matrix to draw the performance map for each aspect, they are able to give a summative rating (Leo & Lachlan-Haché, 2012). The advantages of this approach are given by Leo and Lachlan-Haché (2012) as follows: through this approach, districts or schools can set a minimum efficiency for each aspect of performance in order to meet overall performance expectations and ensure that teachers meet the expectations (thus avoiding misclassified results). Using this approach, it is possible to differentiate between teachers based on their experience, subjects taught, students' grades and by district, thus setting unique

expectations for teachers. This approach is appropriate for providing feedback to teachers by highlighting their strengths and weaknesses since they are not obscured by an averaging process. Through this approach, evaluators can use scale, qualitative or binary data in combination to determine teacher performance. For instance, student achievement data can be rated on a numeric scale of 1 to 4, a qualitative observation rating can be given, such as unsatisfactory, etc., and binary rating can be given to indicate professionalism, such as, does or does not meet expectations. Using this approach, the summative results can also include both qualitative and narrative data that are useful, since some data that is collected using certain tools may be not always translate into a numerical value.

However, there are drawbacks to this approach. As pointed out by Leo and Lachlan-Haché (2012), this approach groups teachers into similar categories, thereby overlooking individual differences between teachers within those categories. Therefore, interpreting the data is more complex when it is used to make decisions on dismissal or promotion. With this approach, evaluators are required to consider different data with equal weight to the rest of the information, which might not be weighted, to determine the final summative score.

Another approach to using multiple tools is *holistic.* According to Leo and Lachlan-Haché (2012) in this approach, evaluators collect data using multiple tools and identify patterns in teacher performance. Then, they compare the data to a similar set of performance criteria. The evaluators interpret the data using the performance criteria to make conclusions about overall performance and to determine teacher performance. This approach is a flexible approach for evaluators since they can take into account a variety of aspects, such as a teacher's knowledge of content and teaching responsibilities, in order to emphasise patterns over any individual data. This approach depends heavily on the evaluator's judgement in determining performance. In this sense, evaluators can play a role in the teacher evaluation system to prevent misclassifying teachers by identifying low-performance teachers. This approach also lends itself well to improving the focus on the implementation of the teacher evaluation system (purposes), feedback aligned with district priorities, the school focus, and individual teachers' targets. This approach requires gathering more information through the use of the tools that are outlined in the teacher evaluation system. However, there are also drawbacks to this approach. First, this approach relies on evaluators being properly trained since it would be difficult to ensure continuity,

consistency and accuracy in passing judgement on teachers across schools and districts if evaluators do not have access to intensive and ongoing training. Second, through this approach, if evaluators do not make plans to provide regular update their data, the results of teacher evaluation may suffer from a lack of transparency.

With regard to the tools that could be used together to evaluate teacher performance, Hanover Research (2012) note that there is no universal agreement about what is best, but that common multiple measures that could be included in teacher evaluation are student achievement data, observation, student reports, portfolios, peer reviews, and parent surveys. Nolan and Hoover (2008) have a similar list, suggesting using administrative observation, peer observation, peer input, teacher portfolios, student evaluation, parental input and student learning data as measures. Some of these tools will be analysed and discussed separately to underline their strengths and weaknesses and to demonstrate how each tool can be made as effective as possible in order to generate the greatest benefits from its use. Thus there is no perfectly right or completely wrong tool; as Goe, Holdheide, and Miller (2014) noted in their practical guide to designing a teacher evaluation system, all tools have their weaknesses and strengths in terms of reliability and validity.

2.7.1 Classroom observation

Through classroom observation, an evaluator can obtain rich information about classroom behaviours and activities, as well as be able to reflect on teaching practice for both formative and summative evaluations. An evaluator can also evaluate the interaction between teacher and students in terms of learning (Goe et al., 2008, 2012; Burnett et al., 2012). Furthermore, through observation, teachers can obtain feedback on their practice more quickly. Whitehurst, Chingos, and Lindquist (2014), in a study of four urban districts in the U.S., concluded that observation is faster than other tools used in teacher evaluation with regard to providing feedback to the teacher for improving performance. Feedback from students' achievement data via standardised tests to reflect teachers' performance, is = often not communicated quickly enough to the teacher.

Moreover, by using classroom observation as a tool for evaluating teachers, both teachers' performance and students' achievement could be improved. Taylor and Tyler's analysis of data from Cincinnati public schools in Ohio (2011) confirmed this result, finding that classroom observation (by external and internal evaluators) improves mid-career teachers'

effectiveness in promoting students' achievement in math, whereas they found no effect on students' achievement on reading tests. They also found that teachers' performance is improved both during the year they are evaluated by classroom observation and the year after.

However, the teacher's contribution or activities outside the classroom cannot be included when observation is the only tool used in teacher evaluation (Goe et al., 2008). Furthermore, evaluators cannot either determine if a teacher's students have achieved growth as expected, collect information to reflect teacher's ability to collaborate with colleagues, or determine if a teacher is communicating with parents effectively (Goe et al., 2014; Goe & Croft, 2009). Also, in some contexts, observation might be expensive due to the cost of training and calibrating to ensure validity, and the cost of the evaluator's time (Goe et al., 2008).

In order to benefit from observation and to ensure validity and reliability, three points should be taken into consideration. Firstly, reliability and validity are improved and enhanced when: frequency of classroom observation is increased; observations occur in different periods of time (different days) (Denner, Miller, Newsome, & Birdsong, 2002; Cronin and Capie, 1986). When observations are infrequent and brief during the school year, they lead to inattention to performance, as indicted by Weisberg, Sexton, Mulhern, and Keeling (2009), who conducted a survey study with approximately 15.000 teachers and 1,300 administrators (evaluators) in 12 districts in Arkansas, Colorado, Illinois, and Ohio. Secondly, the observation should be subject-specific, as Hill and Grossman (2013) indicated, which means that the observation should be related to the school level and subjects. Generic observation is limited in determining teacher performance. Thirdly, conducting observations by different observers for multiple observations increase reliability as the deployment of different observers reduces the likelihood of an unusual judgement and the influence of an atypical lesson, as found by Kane and Staiger (2012) in their report investigated observation as a tool alongside other measures of teaching for the Measured Effective Teaching (MET) project, which analysed 7,491 videos of instruction of four to eight lessons given by 1,333 teachers in grades 4-8 in six districts in the US. The MET project developed measures to reflect all aspects of effective teaching, including student surveys to evaluate the instructional environment, content tests to assess teachers' knowledge of their subjects, observations to evaluate their practice, and student

assessments to reflect the learning outcomes of teacher's learning in order to evaluate alternative ways of providing valid and reliable feedback to the teacher to develop and improve their teaching, but this report focuses on classroom observation. In this report, the results also show that combining observation with student achievement and student surveys improved reliability.

2.7.2 Self-Evaluation

The main benefit of conducting self-evaluation is to help teachers realise the directions that they should follow in their work, and what objectives need to be set for professional development. Self-evaluation can lead to a teacher's increased awareness of their performance, as a result of self-reflection, which may lead them to perceive a need to improve (Zhang, 2008). Therefore, the value of self-evaluation is that it is an opportunity for teachers to determine their own strengths and weaknesses, which may lead them to enhance teaching and learning, as found in Ovando's study that included twelve teachers in primary schools in Texas, and their written responses to open-ended questions (2001).

However, self-evaluation in certain cases can be problematic. A study in two primary schools in Georgia, using documents, artefacts, and interviews to collect teachers' perspectives on teacher evaluation (Looft, 2002), found that the words/language of the self-evaluation form made it difficult for teachers to provide an honest rating. The wording was not always understood and the length of the forms and details that need to be included caused stress to teachers. Moreover, the variation in the level of feelings from day to day may affect the objectivity of self-evaluation, when evaluators require teachers on a particular day to reflect on their teaching over a period of time. Therefore, conducting self-evaluation presents challenges, as misrepresentation and misreporting may have an effect on the report (Goe et al., 2008); in other words, it may not be an accurate reflection of teacher performance. To support the teacher in meeting the challenge to reflect more objectively on what has happened during a lesson, video/audio recordings, peer evaluation feedback, and short stories (written by teacher during the school year) can be used and combined as sources of data (Nikolic, 2002).

Additionally, validity may be a point of concern in teacher evaluation. In Zhang's study (2008), heads of departments and administrators reported that teachers could not translate their performance into realistic scores, and so gave themselves a higher score than their actual performance. Some teachers confirmed that they did over-evaluate to protect

themselves; some tended to overrate their performance, thinking that they did not deserve a low score. On the other hand, some teachers reported that they evaluated their performance accurately and gave themselves high scores simply because they had performed very well.

Although the validity of self-evaluation may be a point of concern in teacher evaluation, it can be improved by comparing it with data obtained through another tool, such as observation in the classroom. As found by Mayer (1999), there was correlation between self-evaluation survey data and classroom observation, enhancing the validity of the results. Eid (2005) tested the validity of self-evaluation in Kuwait by comparing evaluation by heads of departments and those by the teachers in their departments in high schools in five educational districts (64 teachers / 62 heads of departments). The findings showed no statistical difference. With regard to reliability, self-evaluation (survey) can be conducted twice to test consistency. Using this method, Mayer (1999) compared two data sets obtained through self-evaluation and found the results were quite reliable.

2.7.3 Student evaluation

An advantage of student evaluation is that students can offer their opinion about teaching practice in the classroom and teachers' work outside the classroom. For example, Zhang (2008), stated that student evaluation provides information about teaching attitudes and effectiveness, moral education, teachers' performance, classroom management, assessment, and tutorials after class, which is broader in scope because it includes not only what happens inside the classroom but also outside it.

On the other hand, Liu and Teddlie (2005) analysed data from interviews with 18 teachers in six schools in China and reports from academic journal published and found that students, especially young students, often did not realise the purpose of teacher evaluation and did not consider their role as important in evaluating teachers. As confirmed by Wang (2004) and Xu (2004), who conducted their research in a Chinese context, students' evaluation may lead to invalid results; for example, a student may evaluate strict teachers with a lower score, and with a higher score for teachers who are considerate of students (cited in Liu and Teddlie, 2005). In the Kuwaiti context, Eid (2005) compared students' evaluation, teachers' self-evaluation, and heads of departments' evaluation in high schools and found that students tended to see their teachers as better than how teachers saw themselves by self-evaluation or by heads of departments' evaluation. Eid suggested that students may also try to improve their teachers' image in front of evaluators or MOE as a

form of courtesy or out of fear of their teachers. Burr (2015) conducted a study to determine the value of student evaluation from the teachers' perspectives (40 teachers in one Utah high school) and found that when conducting student evaluation in different periods during the school year, teachers reported that students' evaluation caused them to reflect and to inspire changes in their instruction based on students' views. However, teachers also expressed concern and anxiety about how students evaluated them, as well as concern about the reliability and validity of student evaluation as an appropriate tool for rating teachers.

Peterson, Wahlquist, and Bone (2000) analysed 9,765 student questionnaires for teacher evaluation from 27 schools in Utah and found that students of different ages in primary and secondary were able to distinguish between teachers who were able to teach and those that students just liked. They were also able to distinguish between teachers who supported their learning and those who treated students well. This is made possible when the items of the questionnaires are appropriate for the student level. For example, the following items were found to work well with elementary students: "I am able to do the work in class, Teacher is kind and friendly, I learn new things in this class, My teacher is a good teacher, Teacher shows us how to do new things, I know what I am supposed to do in class " (p. 150).

Similarly, Ferguson (2010), in a study of 2358 classrooms, found that students could make valid distinctions about their classroom on seven issues referred to as the Seven C's: *Care* refers to teachers' care for their students, such as taking into consideration students' emotions and reducing their anxiety; *Controls* refers to classroom management; *Clarify* refers to a teacher's role in promoting student understanding and clearing up confusion; *Challenge* refers to teachers' support and encouragement of students to work hard; *Captivate* refers to the teacher's role in making the classroom stimulating and in avoiding making the learning boring; *Confer* refers to teachers keeping students alert in the classroom by asking them about their views and inviting them to express themselves; and finally, *Consolidate* refers to how teachers organise material for students through reviewing and summarising. In short, these seven headings reflect how teachers can teach well and how much students can learn. In terms of the reliability of students' evaluation, Zhang (2008) suggested increasing the number of students involved in evaluating teachers.

In short, though students' feelings about teaching and teachers are valued (Zhang, 2008), student evaluation should be used cautiously, as students lack subject knowledge and teaching experience. Zhang suggested using student evaluation as a reference for teacher performance while Mertler's study (2007) concludes that student evaluation as feedback for the teacher regarding his/her performance is very useful and something that can be for teachers rather than against them. Burr (2015) confirmed that students' feedback could reduce anxiety among teachers, as teachers appreciated the feedback from their students as a means of development, rather than it is being used as a judgement about them. Burr suggested that for feedback from students to be effective, student evaluation should be conducted in two different periods, early in the school year or in the middle and then at the end of the school year, for example, October and March. Timing is also important: schools should avoid conducting it after a major test, as it will affect their evaluation of the teacher (Olatoye & Aanu, 2011).

2.7.4 Peer evaluation

As described above, the teacher can participate in teacher evaluation either as self- or peer-evaluation. Eri (2014) describes his experiences with peer evaluation via observation; he believes that peer evaluation leads to improving teaching regardless of how experienced the observed teacher is. He also believes it is a good exercise for teachers to write feedback. Salih (2013) used a questionnaire to explore teachers' reactions to peer evaluation with 40 English teachers in two higher institutes in Oman. He found that peer evaluation supported the review of teaching, as the teachers' perception in this study with regard to peer evaluation was positive. Teachers believed that the peer evaluation was useful, as it allowed them to exchange feedback with each other, reflect on their teaching, enabled them to offer suggestions to each other, helped them to modify their teaching, and gave them more confidence in their teaching.

On the other hand, using peer evaluation for summative purposes may be unacceptable in some contexts, as found by Joshua et al. (2006) who conducted a study in Nigerian secondary schools comprising 480 teachers, by using a questionnaire. The researchers found the general attitude of teachers in this study towards using peer evaluation for summative purposes was negative, as teachers did not trust them to make decisions about their promotion and rewards.

Similarly, while peer evaluation is noted as helpful for teacher development due to teachers receiving comments and suggestions from each other and sharing experiences, knowledge, and understanding, it is not helpful in all cases. For example, conflicts between teachers may prevent the exchange of reliable and frank comments about colleagues' teaching, and thus peer evaluation loses its function of promoting professional development. When peer evaluation is used to make judgements or give scores, personal relationships might affect the evaluation (Zhang, 2008). Therefore, to be more effective, teachers should be critical friends who conduct peer evaluation without any subjectivity or bias (i.e. teacher should not ignore good parts of teaching and highlight only the weaknesses) as suggested by Salih (2013).

The question raised here is how to make peer evaluation as effective as possible in order to get maximum benefit from it. First, courses should be organised for teachers in order to improve their knowledge and skills and to avoid misunderstanding over why they should engage in peer evaluation (Arnodah, 2013). Second, clear performance guidelines and explicit teaching criteria are key components for peer evaluation to work well (Johnson & Fiarman, 2012). Third, time allocation for conducting peer evaluation has been shown to have a positive effect, as found by Brix, Grainger, and Hill (2014) in a case study of a regional secondary school in Australia.

2.7.5 Portfolio

Some of the advantages of using portfolio can be seen in previous research in that they can provide accurate and comprehensive information about teachers' performance (Attinello, Lare, & Waters 2006) and reflect a teacher's commitment to the teaching profession (Westhuizen & Smith, 2000). In this sense, it provides more comprehensive information about a teacher's performance than observation inside the classroom as a tool in teacher evaluation, as found by Attinello et al. (2006). Moreover, a portfolio as a tool is useful in encouraging teachers to reflect on their teaching (Attinello et al., 2006), and help them identify their personal strengths and weaknesses (Attinello et al., 2006; Dinham & Scott, 2003), thus laying the groundwork for professional development (Attinello et al., 2006; Chorrojprasert, 2005). Furthermore, a portfolio allows teachers to be more collaborative through sharing and discussion with others, as well as allowing them to show evaluators their achievement (Dinham & Scott, 2003; Attinello et al., 2006).

In a study about the perspectives of teachers and administrators in a rural/suburban school district in the south eastern region of the US about using portfolios, Attinello et al. (2006) gave a survey to 23 schools (752 teachers 46 administrators responded). Both administrators and teachers had positive views about using portfolio, but administrators were significantly more positive and supportive of using portfolio as a comprehensive measure and support for teacher's self-reflection, believing portfolios had a positive effect on professional development. Chorrojprasert (2005) conducted a study in a secondary school in Bangkok that included the views of 388 teachers via survey and 9 teachers via interview. The researcher found that almost half the teachers in the study viewed portfolio as an appropriate and efficient tool for both determining performance and professional development. Teachers mentioned responsibilities and activities, official documents, students' progress and work, and personal qualities as all contributing to giving an accurate reflection of teachers' performance. Teachers reported that portfolios also helped them to reflect on their performance and that the process of preparing their portfolio enabled them to plan improvements and implement their plans; they also reported becoming more aware of their students' needs. In one case study conducted by Westhuizen and Smith (2000), teachers stated that they wanted to present their performance and professional skills when being evaluated, and that they viewed a portfolio as being able to reflect their performance and skills.

On the other hand, a number of drawbacks with using a portfolio as a tool for teacher evaluation have been reported. Firstly, the time required to prepare a portfolio is seen as a disadvantage (Westhuizen & Smith, 2000; Attinello et al., 2006; Dinham & Scott, 2003), with some teachers preferring to spend time on preparing lessons or organising activities (Attinello et al., 2006). Secondly, time is also needed for the evaluator to review the portfolio, and as found by Attinello et al. (2006), some administrators did not spend enough time reviewing the portfolio, reporting the task as very time-consuming. Thirdly, portfolios can be daunting for teachers who have no experience with portfolio construction (Dinham & Scott, 2003). Fourthly, a portfolio does not always reflect all aspects of teaching, as noted by teachers and administrators in Attinello et al. (2006): an outstanding teacher may not necessarily create a good portfolio.

Consequently, Attinello et al. (2006) offer some recommendations for using a portfolio: there should be clear guidelines for how to use the portfolio, on-going training for both

evaluators and teachers with regard its use, sufficient time for evaluators to review teachers' portfolio, and it should be used as part of teacher evaluation together with other tools such as observation.

2.7.6 Student achievement data

Student achievement data are used in teacher evaluation to evaluate teaching (Hanover Research, 2012). Darling-Hammond et al. (2012a) suggest that one aim of using student achievement information is to make student learning a part of teacher evaluation.

There are benefits to using student achievement data in teacher evaluation. Goe and Croft (2009) argue that evaluators can directly focus on and analyse student learning to determine a teacher's contribution. An evaluator can also compare student achievement data across classrooms, schools, and districts to make judgements on the teacher's learning outcomes (Burnett et al., 2012). However, Goe and Croft (2009) point out that while student achievement data could be helpful for evaluators when determining a teacher's contribution, this will not give weak teachers the information they need to help them improve their performance. Furthermore, standardised testing can be expensive in some contexts, especially in terms of conducting the test in all the districts, designing or purchasing the tests (Burnett et al., 2012), and hiring experts to analyse the results (Goe & Croft, 2009). In addition, Baker et al. (2010) and Hanover Research (2012) point out potential negative consequences of using student achievement data to determine sanctions and rewards for teachers, since relying on test score results can dissuade teachers from working with high-needs students and discourage outstanding teachers from teaching classes with a large number of high-need students (who are traditionally 'weak').

There is enough widespread discussion of the negative effect of testing on both teachers and students that it deserves some attention. Smith and Rottenberg (1991) indicated that standardised tests might cause students to experiences stress. Jones, Jones and Hargrove, (2000) also claim that high-stakes testing may induce stress for students. According to Stecher (2002, p. 86) the negative effects on students of high-stakes testing is that tests "frustrate students and discourage them from trying, making students more competitive, and cause student to devalue grades and school assessments". Using student achievement data via standardised tests may cause teachers to teach students to the test skills (Hanover Research, 2012; Jones et al., 2000). Teachers coaching students for tests can also have

negative effects, including narrowing of the curriculum to those aspects which are tested. As indicated by Stecher (2002) and Jones et al. (2000), testing students with high stakes forces teachers to focus more on the specific content of the test than on other aspects of the curriculum. The potential effect of high-stakes testing on teachers also might tempt teachers to cheat when administering tests (Stecher, 2002).

There are also concerns regarding using student achievement data as a measurement of teacher effectiveness, according to Darling-Hammond et al. (2012a) and Schafer et al. (2012), since teacher effectiveness varies from class to class, across different grades and from one statistical model to another. Moreover, some teachers have many students in their classrooms with poor attendance, which may affect their level of achievement (Darling-Hammond et al., 2012a). Furthermore, there are differences between teachers with regard to which elements of teaching effectiveness relate to positive/negative student achievement, since some teachers have a greater impact on some parts of student learning than on other parts. Therefore, determining what is relatively more or less effective depends on the tests that are used (Darling-Hammond et al., 2012a).

There are also factors affecting the validity of student achievement results. Darling-Hammond et al. (2012a), Darling-Hammond, Cook, Jaquith, and Hamilton (2012b), and Baker et al. (2010) found that student learning is not influenced by teaching alone. Firstly, there is the school factor, consisting of curriculum, class size, resources, teaching time, and available teaching material. Secondly, students' home lives and communities, as well as family income levels, may have an effect on their academic achievements. Thirdly, the achievements of students with special educational needs and abilities may not provide an accurate reflection of their teachers' efforts. For example, teachers appear more effective when teaching very good students than they are when teaching students with special education needs. Fourthly, a student's peers can have both positive and negative effects on student achievement. Lastly, previous teachers, schools, and other current teachers may also influence student achievement.

Consequently, Baker et al. (2010) argue that student test scores should only be used as a minor element in a broader set of evidence indicating teacher practice. They also point out that statisticians, psychometricians, and economists agree that student achievement data alone are not reliable nor sufficiently robust when making decisions on sanctions and rewards. Similarly, Newton, Darling-Hammond, Haertel, and Thomas (2010) state that

student achievement data should be used with caution when evaluating teachers, especially when making decisions about them. Darling-Hammond et al. (2012b) suggest that student achievement calculated via standardised tests can be included, but that the focus should be on the relationship of the tests to the curriculum and what the students are being taught.

2.8 Giving appropriate feedback to teachers

According to Delvaux et al (2013), the nature of feedback that teachers receive is an important feature of a teacher evaluation system. The literature lists a range of potential benefits. It may help teachers think critically about their teaching practice (Donaldson & Peske, 2010); regular and specific feedback to teachers may give clear information about their strengths and weaknesses, and pinpoint areas that need improvement with recommendations on how to improve their teaching, as some teachers have shown in the SCORE report (2012); feedback may encourage teachers to collaborate with each other, as feedback leads teachers to have conversations about their performance with colleagues and to help each other, especially with experienced teachers and newer teachers, as found by the SCORE report (2012).

To derive maximum benefit, feedback should be provided to teachers immediately. Scheeler, Ruhl, and McAfee (2004) conducted a systematic search of empirical literature, and found that only immediate feedback was considered effective, as it prevents the teacher from continuing to make errors in his/her teaching, while with delayed feedback, the teacher may continue to make errors instead of changing or improving. Khachatryan (2015) suggest that teachers learn best from feedback that consists of specific comments and detailed recommendations. In a single-case study with an administrator and four teachers as participants, Khachatryan found that feedback sometimes did not provide a sufficiently clear picture to enable the teacher to improve. In addition, feedback can be given orally or in writing and evaluators should consider which type is appropriate for teachers' needs. One example is a conversation before and after classroom observation. Such feedback conversations between teachers and evaluators lead to increased levels of trust and collaboration between each other (SCORE, 2012). A second example is written feedback, which allows teachers time to read, interpret, review, and internalize (Kelly, 2014). Furthermore, evaluators should take into account teachers' experience when providing feedback to teachers so as to be appropriate for each individual performance. Tuytens and Devos (2012) conducted a study via interview with school leaders and

questionnaires with 298 teachers in 32 Flanders schools. They found that teachers with more experience see feedback as less useful than teachers who are less experienced. Delvaux et al. (2013) also found that among teachers with limited teaching experience, feedback is positively related to the effect of teacher evaluation on professional development.

2.9 The importance of qualified evaluators

Evaluators play a key role in teacher evaluation. Therefore, there is some concern regarding the evaluators' role and qualifications and how these impact on a teacher evaluation system.

First, lack of background in teaching and subject knowledge of evaluators may lead to invalid and unreliable reports about teacher performance, as found by Albustami (2014), who conducted qualitative research including 5 supervisors, 5 principals, and 10 teachers in Abu-Dhabi schools. The knowledge of evaluators about the subject, pedagogies, and experiences of teaching also allow evaluators to identify what teachers have done, and to anticipate what teachers need to assist them with their development and improvement (Donaldson & Peske, 2010). Accordingly, evaluators should have experience in teaching and subject knowledge to be able to determine teachers' performance and to identify what requires improvement and further development. On the other hand, although the evaluator should have content expertise, meaning that he or she should have an understanding of the content or an understanding of how students encounter the content, content expertise is not always required to evaluate elements of teaching such as managing behaviour, motivating students, building learning environments, as these are common across subjects (Hill & Grossman, 2013).

Second, even if evaluators agreed on their judgement about teachers, their decision is not always valid. A reason for this is that evaluators sometimes are not trained well to conduct teacher evaluation or have little experience in evaluating teachers, as found by Albustami (2014). Accordingly, training is essential (Darling-Hommand et al., 2012b; Albustami, 2014; Partee, 2012) in order for evaluators to be able to evaluate teachers effectively, and thus make the overall teacher evaluation system reliable (Albustami, 2014; Nolan and Hoover, 2008). Training courses for evaluators should include instruction on how to evaluate (Donaldson & Peske, 2010), provide beneficial feedback, provide on-going

support to teachers (Darling-Hammond et al., 2012b), and collect evidence and evaluation reporting (Partee, 2012), as well as skills for analysing effective teaching practice (Albustami, 2014).

Third, the relationship between evaluators and teachers may affect teacher evaluation. Delvaux et al. (2013) found that if the relationship between evaluators and teachers is determined as too positive, the effect of teacher evaluation on professional development may be weaker, because teachers may feel less pressured to undertake actions. Furthermore, Zhang (2008) found that teachers who obtained low ratings on their performance seemed unwilling to improve their performance, as they saw the judgement negatively due to their belief that other teachers obtained a high rating because they had a positive personal relationship with the evaluator/s. Accordingly, an evaluator should control their personal relationship and feelings with teachers, in order to guarantee that the relationship does not affect the evaluators' role, and thus the teacher evaluation.

2.10 Previous studies on the teacher evaluation system in Kuwait

In this section, previous studies on the teacher evaluation system in Kuwait are analysed. There are few studies about the Kuwaiti teacher evaluation system, according to the researcher's knowledge and search. The previous studies will be reviewed in two sub-sections starting with the previous system (no.461/1993) then moving on to the current teacher evaluation system which replaced it in 2006 (no.36/2006). While in the previous system, the teacher was evaluated by the head teacher and inspectors, in the current system, as explained before, the teacher is also evaluated by the head of department. In the current system, teachers are not informed about their evaluation reports (MOE, 2011; KTS, 2010), while previously, teachers were informed about their evaluation report in the middle of the school year but the final report at the end of school was kept confidential (Alkhayat & Dhiab 1996; Alhamdan, 1998). The criteria of both systems, however, are somewhat similar (further details about these systems, see Appendix 1).

2.10.1 Teacher evaluation 461/1993

Two studies have evaluated this system. Alkhayat and Dhiab (1996) conducted research by using questionnaires and a sample of teachers, head teachers, and inspectors in primary, middle, and high school in different educational districts. The researchers found that the system promoted professional development. For example, it contributed to preparing

training courses for teachers' needs, and showed the strengths and weaknesses of teachers' performance. The system determined teacher performance according to the extent to which educational targets were achieved and the extent to which teachers were able to teach (Alkhayat and Dhiab, 1996). Also, the system achieved the purpose of teacher evaluation related to sanctions and rewards, such as teachers' promotions or salary increase. The criteria used to evaluate the teachers were found to be appropriate, as participants had positive attitudes towards them.

From a sample of 406 teachers, 50 head teachers, and 104 inspectors at the high school level in five educational districts (Asimah, Ahmadi, Jahra, Farwaniya, Hawalli) Alhamdan (1998) concluded that some of the criteria used to evaluate teacher performance in the previous system required clarification as evaluators and teachers interpreted them differently, and thus were difficult to measure. The researcher also found a mid-year report helped teachers to improve their performance in the second half of the term. However, giving teachers a score rating in the mid-year report caused problems between teachers and head teachers when a teacher obtained an unexpected score. Therefore, the researcher suggested that the mid-year report should only include strengths and weaknesses without a score.

With regard to the final report of summative evaluation of individual performance, Alhamdan found that teachers' opinions varied. Some supported keeping the reports confidential because they could cause problems among teachers and between teachers and evaluators (head teachers and inspectors). Others supported informing teachers about their strengths and weaknesses without including a score of their performance, except for teachers who were underperforming.

The researcher provided four main recommendations. Firstly, he suggested the scoring be changed to include "very good" with "outstanding, good, weak". Second, he recommended that the head of a department participate in teacher evaluation, as he or she views teacher practice and activities more than any other kind of evaluator. Third, teacher evaluation should focus on teaching activities more than non-teaching duties.

2.10.2 Teacher evaluation 36/2006

Two studies were also found about the current system. A study by Sabti (2010) indicated some drawbacks of the system. First, there seems to be a lack of appropriate training and workshops to improve teachers' performance. Such training, he claimed, should be based on the reports of the teacher evaluation, but this is rare. Second, teachers have no role in suggesting training courses and workshops. Third, evaluation depends solely on observation by head teachers, inspectors, and head of department. In this study, the researcher recommended that teachers should attend a training course or workshop every year or every two years and that evaluators should use a range of tools for teacher evaluation.

The second study (Alsanafi, 2012) involved a sample of 110 social science teachers in the middle level school in two educational districts (Asimah and Mubarak Al-Kabeer). The research concluded that the system was largely successful in promoting professional development and determining teacher performance. However, teachers did not obtain monthly feedback from evaluators and the system is not appropriate for making decisions about sanctions and rewards.

Other drawbacks in the current system are highlighted in this study: firstly, teacher evaluation was found to be too subjective; second, individual teacher evaluation reports were kept confidential, in other words, neither the mid-year report nor the end-of-year evaluation report were made available to the teachers. The researcher also found that while criteria overall are appropriate for social science teachers, but that some criteria needed more clarification in order to generate a consistent interpretation.

Alsanafi (2012) concluded that teachers should be given a detailed mid-year report, and the final report of the evaluation should also be provided to teachers but without a grade (only comments about teacher's performance) to avoid causing problems between teachers and head teachers. She also recommended that teachers should be more involved in teacher evaluation and work together with inspectors, head teachers, and heads of departments.

2.11 Summary

This literature review has been based on the premise that an effective teacher evaluation system is one that determines teacher performance accurately, supports decisions in relation to sanctions and rewards, and promotes professional development. It has analysed the factors that contribute to teacher evaluation based on findings in previous studies. Several factors have been highlighted as of particular importance. First, a teacher evaluation system has to comprise both professional development and summative evaluations as each purpose serves different needs. Second, the criteria for evaluating teacher performance should be explicit in order to generate agreement of both evaluators and teachers of what constitutes effective teaching, to promote better understanding of effective teaching, and hence to inspire better performance. Third, external and internal evaluators should be involved in evaluating teachers, as each type offers benefits. Fourth, a teacher evaluation system should include multiple tools, as there is no perfectly right or completely wrong way and the use of various tools helps make a teacher evaluation system effective. Fifth, an effective system provides feedback (immediate feedback, the specific feedback, and type of feedback) and has qualified evaluators (the background of evaluators, training courses for evaluators, and controlling personal relationships between evaluators and teachers).

Chapter Three: Research Methodology

3.1 Introduction

This chapter presents the design of the study. The study used a mixed method design, and therefore starts by discussing and explaining both qualitative and quantitative research. Thereafter, it describes the instruments of data collection that were used, namely, questionnaire, interview, focus group. Next, it justifies the subject population and sample that were chosen. Finally, it explains the data analysis and assesses the quality of the data.

For Creswell (2012, p.3) research is "a process of steps used to collect and analyse information in order to increase our understanding of a topic or issue. At general level, research consists of three steps: pose a question, collect data to answer the question and present an answer to the question". Some research consists of more than three steps but these steps are the core elements of any research (Creswell, 2012). Accordingly, research is one of several different ways to obtain understanding and knowing (Mertens, 2015).

There are many reasons to conduct research in the field of education. Perhaps the broad reasons are *to explore, to shape policy,* and *to improve practice.* To explore means that researchers intend to include everything from finding answers to research question, to identifying particular problem or issues that should be the subject of further research. To shape policy refers to when researchers intend to collect information in order to make judgments about whether policy targets have been achieved or are at least going in the right direction. To improve practice is when researchers conduct research in order to provide suggestions for the reform of that which has already been done, so as to improve performance, institutions, education outcomes, personal effectiveness as teachers (Newby, 2014). In this context, it can be stated that this research aims to *shape policy* by informing policy makers about how teacher evaluation is working and to present to them the perspectives of teachers, inspectors, and head teachers regarding the current teacher evaluation system and what they would like teacher evaluation to look like. Moreover, this research aims to *improve practice* by suggesting an alternative system for evaluating teachers in order to make better use of teacher evaluation in Kuwait.

In terms of research type, this research is descriptive. Best and Kahn (2006, p.118) define descriptive research as research which:

> …describes and interprets what is. It is concerned with conditions or relationships that exist, opinions that are held, processes that are going on, effects that are evident or trends that are developing. It is primarily concerned with the present, although it often considers past events and influences as they relate to current conditions

Descriptive research was selected because this researcher wanted to describe and interpret what actually happens in Kuwait and discuss what should happen in teacher evaluation. As noted by Cohen, Manion, and Morrison (2007), descriptive research is used to describe, compare, determine differences, classify, analyse and interpret various events of inquiry by looking at individuals, groups, schools or institutions and materials. Descriptive research can be also used as a basis for suggesting answers to questions, e.g., how something should improve and what the best way is to do so, what the reactions of the participants are (Knupfer & McLellan, 1996). Moreover, this research also analyses the extent to which the proposed alternative system for teacher evaluation is appropriate and workable in the Kuwaiti context.

Finally, it should be noted that this approach was selected as the best fit for answering research questions, as the MOE and CSC use the results of teacher evaluation to support their laws and regulations governing education (see Appendix 1). Therefore, it was difficult to conduct, for example, experiment study, and the time horizon would be at least four years for applying alternative systems.

3.2 The nature of the research

In order to properly answer the questions, this research used a combination of quantitative and qualitative methods for data gathering. According to Bryman (2006), a combination of qualitative and quantitative research is referred to as a mixed method design. In the following three sub-sections, qualitative and quantitative research are discussed, before light is shed on mixed methods design.

3.2.1 Qualitative research

By adopting a qualitative research approach, researchers seek to acquire rich details of the topic being studied (Creswell, 2012; Johnson & Christensen, 2008). Perhaps the most important reason to acquire rich details is so that, as Newby (2014) argues, researchers are not limited by numerical considerations when seeking to create understanding and finding answers from any evidence that reflects motives, values and attitudes. This evidence is collected through what participants and researchers say, what they do, pictures, the writing they produce, and the objects they create which are then evaluated and interpreted (Newby, 2014; Check & Schutt, 2012). Secondly, this approach makes it possible for researchers to ask broad questions; therefore, open-ended questions can be posed (Creswell, 2012; Mack, Woodsong, MacQueen, Guest, & Namey, 2005). Thirdly, researchers can benefit from the flexibility that is inherent to this approach, such as modifying the procedures during the research period, refining the focus to change or develop a deeper understanding of the context or discovering new aspects of the area under discussion (Johnson & Christensen, 2008; Mack et al., 2005).

By asking open-ended questions, remaining flexible and collecting a variety of evidence based on what the participants do or say, this researcher was able to demonstrate a variety of perspectives and reveal the participant's knowledge and practice, as well as take into account his/her social background. All of this falls under the category of qualitative research, as noted by Flick (2009). Thereby, this researcher could create a better understanding of an individual's experiences of a particular topic, as noted by Johnson and Christensen (2008).

However, there are some weaknesses in qualitative research. First, the results might be unique to the people that are included in the study. Second, despite the lengthy amount of time that is often needed to complete the data analysis, it is very difficult to make quantitative predictions. Third, the results might be influenced by personal bias and idiosyncrasies. Fourth, the research might be accorded lower credibility by some program administrators and representatives (Johnson & Christensen, 2008).

3.2.2 Quantitative research

Descriptive quantitative research aims to discover the frequency and distribution of the topic that is under investigation (Flick, 2009). It can also analyse trends, discover or

explain a relationship, compare variables, or identify differences and similarities between groups (Mertens, 2015; Creswell, 2012). By adopting quantitative research, researchers can also use established methods in order to deal with, or discuss, numerical data (Gorard, 2008; Johnson & Christensen, 2008; Newby, 2014). These methods are structured and can include questionnaires or structured observation (Mack et al., 2005).

In adopting a quantitative research approach, the research has to be planned from start to finish so researchers cannot react to the participants' responses. This is in contrast to the qualitative research approach in which researchers can determine how and in what order the questions are posed. However, through quantitative research, researchers can collect data from a large number of participants and analysis is less time-consuming (Johnson & Christensen, 2008). Consequently, in adopting quantitative research, this researcher was able to discover the frequency and distribution of the responses of a large number of teachers to gain their perspectives on the current teacher evaluation system and assess whether differences exist between teachers with regard to their gender, experiences, subjects taught and educational districts.

In terms of the weaknesses associated with quantitative research, firstly, the categories used by researchers might not reflect the local constituencies' understanding. Secondly, the knowledge produced might be too abstract and general when attempting to apply it directly to specific contexts, situations or individuals. Thirdly, researchers might miss certain phenomena because of the heavily weighted focus on theory or hypothesis *testing* as opposed to theory or hypothesis *generation*. Fourthly, the specific theories outlined by researchers may not necessarily reflect the understanding of local constituencies (Johnson & Christensen, 2008).

3.2.3 Mixed method designs

Mixed method can be defined as the combination of both quantitative and qualitative research to use in a single study to answer the research questions. This can reflect a combination of different elements pertaining to data collection, analysis, integration of the results, and deduction techniques (Tashakkori & Creswell, 2007; Johnson, Onwuegbuzie, & Turner 2007).

There are different purposes for using a mixed method design. Greene, Caracelli, and Graham (1989, p.259) suggest that these purposes can be divided into five categories: triangulation, complementarity, development, initiation and expansion (see Table 3.1).

Table 3.1: Purposes for mixed method evaluation designs

	Purpose	Rationale
Triangulation	Seeks convergence, corroboration, correspondence of results from the different methods.	To increase the validity of constructs and inquiry results by counteracting or maximising the heterogeneity of irrelevant sources of variance attributable especially to inherent method bias but also to inquirer bias, bias of substantive theory, biases of inquiry of context.
Complementarity	Seeks elaboration, enhancement, illustration, clarification of results from one with the results from the other method.	To increase the interpretability, meaningfulness, and validity of constructs and inquiry results by both capitalising on inherent method strengths and counteracting inherent biases in methods and other sources.
Development	Seeks to use the results from one method to help develop or inform the other method, where development is broadly construed to include sampling and implementation, as well as measurement decisions.	To increase the validity of constructs and inquiry results by capitalising on inherent method strengths.
Initiation	Seeks the discovery of paradox and contradiction, new perspectives of frameworks, the recasting of questions or results from the other method.	To increase the breadth and depth of inquiry results and interpretations by analysing them from the different perspectives of different methods and paradigms.
Expansion	Seeks to extend the breadth and range of inquiry by using different methods for different inquiry components.	To increase the scope of inquiry by selecting the methods most appropriate for multiple inquiry components.

- Greene et al. (1989, p.259).

The main purpose of using mixed method in the current study is expansion, wherein this researcher extended the range of inquiry by using different methods for data collection and data analysis. A questionnaire was used for the collection of data about teachers' perspectives on the current teacher evaluation system. Interviews with head teachers and inspectors were conducted to collect data about the current teacher evaluation system and to test an alternative system; furthermore, a focus group composed of teachers was used to gather data on a potential alternative system.

A secondary aim of using mixed method in the current study is development. This researcher used the results taken from the questionnaires and interviews that reflected the participants' perspectives on the current teacher evaluation system in order to inform and help draft some of the points that were used when proposing an alternative system (see Chapter Six, Section 6.4).

Moreover, it is apparent from the evaluation of quantitative and qualitative research presented above that there is no right or wrong way to approach data collection; the combination of qualitative and quantitative methods might help to achieve better results. As Biesta (2012), by using mixed methods, this researcher tried to design a study that benefited from the strengths of both the two approaches. This researcher hoped that the weaknesses of one approach would be supported or counterbalanced by the strengths of the other.

This research took place in three stages, with each stage including several different methods of data gathering:

In the first stage, this researcher collected data perspectives to answer RQs 1, 2, 3 and 4 by giving a questionnaire to teachers and interviewing inspectors and head teachers.

In the second stage, this researcher proposed an alternative system to head teachers and inspectors in order to ascertain their views (RQ5). The system was introduced through written materials that were read by the participants beforehand and by data gathering carried out during the interviews.

In the third stage, this researcher proposed the alternative system to the teachers in order to elicit their views (RQ5). Teachers were introduced to the system through written materials, which were also described in the focus group interviews, and this served as a prompt for the data gathering.

3.3 Data gathering instruments

As mentioned in the previous section, methods used were questionnaires in combination with individual and focus group interviews.

3.3.1 Questionnaire

Brown (2001, p.6) define questionnaires as "any written instruments that present respondents with a series of questions and statements to which they are to react either by writing out their answers or selecting from among existing answers". Although questionnaires are basically used to collect information, they can have a variety of purposes. Tymms (2012) lists four purposes: exploratory work, describing a population, outcomes or controls in studies, and feedback. These are described below.

- Exploratory work: when researchers intend to investigate a phenomena but are unsure of the best way to proceed, questions can be asked of individuals taken from the target populations or colleagues can be consulted and literature read to shape what exactly is observed. Afterwards, the researcher may distribute questionnaires to collect data from a small sample to help to define a problem to follow up by more in depth methods.
- Describing a population: when researchers would like to identify a general pattern across a population through administering questionnaires to a representative sample of the population.
- Outcomes or controls in studies: when researchers intend to conduct a questionnaire as part of an intervention study or quasi-experiment, which needs its results to be measured and compared. In such a case, questionnaires are used in order to assess somebody's understanding and knowledge in a manner that is similar to the purpose of testing but with less pressure and more diversity.
- Feedback: in this case, the questionnaires are used in in-service, courses, or during training in order to obtain information on the attendees' experiences. Here, the purpose could be formative in order to improve for the next time, or the results could be used for summatively.

In this research, a questionnaire was used to collect teachers' perspectives on the current teacher evaluation system and their views on what should dominate teacher evaluation. Therefore, the questionnaire was designed to obtain feedback and describe a population, as outlined by Tymms (2012) above.

There are some inherent advantages to using a questionnaire that compelled this researcher to use one in this study. Questionnaires are used in order to obtain a large volume of data; they can consist of several questions and be administered to a large sample (Mertens, 2015). Where a large sample can be gathered, questionnaires can be conducted quickly and easily (Burton & Bartlett, 2009) and at a low cost (Mertens, 2015). Moreover, questionnaire respondents might feel more comfortable when giving their responses due to the anonymity of the questionnaire and lack of face to face contact with the researcher (Cohen et al., 2007). Finally, the data collected from questionnaires are suitable for, and easy to use in, analysis and in comparison with other results (Burton & Bartlett, 2009).

However, one drawback with regard to questionnaires is that some participants may not complete them in their entirety or may not return them at all (Cohen et al., 2007). To avoid that, this researcher excluded sensitive questions from the questionnaire and ensured that it was not too long. This researcher also administered the questionnaire through the MOE. As pointed out by Edwards et al. (2002), questionnaires that are administered through organisations such as universities can positively affect the number of returned questionnaires since participants are more likely to return them to public bodies than they are to other sources, such as commercial organisations. Another issue is that some participants might not give serious attention to answering the questions or they may misunderstand some of the words in a questionnaire (Mertens, 2015). This researcher attempted to avoid some of these problems by providing explanations for any terminology that could be misconstrued.

3.3.1.1 Construction of the questionnaire

Questionnaires can consist of different types of questions, either open-ended or closed (Cohen et al., 2007; Tymms, 2012; Check & Schutt, 2012; Newby, 2014). Open-ended questions are not limited to a set list of choices and the participants write their responses in their own words (Check & Schutt, 2012). Closed questions are restricted to a list of responses from which the participant can choose. These are mainly Likert-type responses. A Likert-type question "involves presenting answers on a scale where the number of possible responses can vary from three up to seven or more" (Tymms, 2012, p.233).

The questionnaire in this study was designed with closed questions, as they are useful for generating frequencies of response, enabling comparison between groups, and aiding participants to be direct and to the point (Cohen et al., 2007).

Before constructing the questionnaire, this researcher reviewed the relevant literature pertaining to teacher evaluation; however, the actual questionnaire was unique to the current study. The questionnaire was designed as follows: a cover sheet, some background questions and four sections with Likert-type questions. The cover sheet included the title and purpose of the study as well as the details of this researcher (name, contact, and name of programme), as recommended by Cohen et al. (2007). Then, background questions were asked about the participant's gender, experience, educational districts and subjects taught, following Newby's suggestion (2014) that these questions are better at the beginning of a questionnaire. These background questions were necessary to identify differences between male and female, their experiences in teaching, and the subjects taught with regard to teacher evaluation. The background question about educational district was meant to determine if the size of a district and the number of schools and teacher in that district make any difference in teacher evaluation.

The four sections using Likert-type questions consisted of the following (a complete version of the questionnaire can be found in Appendix 2 for English, and Appendix 3 for Arabic):

- Section one asked teacher participants to determine the actual purposes of the current system and compare this to their desired purposes. In this section, three purposes of teacher evaluation were given, and each purpose was rated according to "frequency" and "importance" (see Table 3.2). The questions were analysed individually and not in an aggregated scale.

Table 3.2: An example of a two-sided question in the questionnaire

Statement	**How often are these used?**						**How important are these?**				
	Never	Seldom	Sometimes	Often	Always		Not Important at all	Unimportant	Neither Important nor unimportant	Important	Very Important
Promoting professional development of teachers			X							X	

- Section two had six items asking teachers about the tools of teacher evaluation. This section intended to identify tools that are used in the current system and to compare this to the tools that should be used. This section also had two sets of

questions for each tool, asking (A) to what degree the tools are used, and (B) to what degree the tools should be used. Again, each item was analysed separately not in an aggregated scale.

- Section three listed three sets of items about each evaluator (inspector, head teacher, head of department). This section intended to measure the involvement of the evaluators in the current system. Each set had questions asking teachers to assess (A) the role of the evaluators and (B) rate the value of their role. These questions were designed to create an aggregated scale to cover the range of teachers' views about the role of evaluators in providing written feedback, engaging in discussion before and after teaching observations, and rating the value of evaluators' role.
- Section four had 14 items that this researcher intended to use to make an aggregated scale to measure the extent to which the current teacher evaluation system supports teachers. These scales included the follow items:

  A) Six items to measure the extent to which the current system supports teaching development. For example, the following items are measured: the support of the system regarding better use of pedagogies, clearer understanding of lesson planning, and clearer understanding of what constitutes effective teaching.

  B) Six items to measure the extent to which the current system supports learning improvement. For example, teachers' abilities to provide students with effective feedback, dealing with individual differences between students, and dealing with students' disciplinary and behavioural problems.

  C) Two items to measure the extent to which the current system supports the awarding of promotions, and rewards (e.g., annual bonuses or salary increases, promotions, etc.).

3.3.1.2 Translation of the questionnaire

The official and native language in the Kuwaiti context is Arabic, but the questionnaire was designed in English. Hence, the questionnaire had to be translated for the participants into their mother tongue. The questionnaire was first translated by this researcher, and it was checked, English to Arabic, and Arabic to English, by two assistant professors at Kuwait University and in the School of Basic Education (Public Authority for Applied Education and Training, PAAET) in Kuwait who specialise in translation between English

and Arabic. Afterwards, this researcher asked an English teacher with 20 years' experience in Kuwait to check it over as well.

3.3.1.3 Piloting the questionnaire

Piloting means trialling the questionnaire in order to increase its reliability and validity (Cohen et al., 2007) through making changes based on feedback obtain from individuals who complete and evaluate the questionnaire (Creswell, 2012). According to Newby (2014), the pilot study should start with the first draft of the questionnaire, which should be given to some experts in questionnaire design in order to revise and modify it. Then, researchers should ask some participants to complete the questionnaire in order to obtain their feedback. These participants' answers are not be used in the study, so researchers should find people with similar characteristics to the population but who will not be part of the study.

This researcher conducted a pilot study as part of the present study for several reasons: to check the clarity of the questionnaire and its items; to reduce difficult and ambiguous words in the items or decide whether to add explanations; to obtain feedback from a sample on the questions and the format, such as rating scales, multiple choice, and so on; to check the time needed to complete the questionnaire; to establish whether the questionnaire is too long, too short, or too difficult; and finally to discover commonly misunderstood or incomplete answers across the pilot participants' responses (Cohen et al., 2007).

PhD supervisors at Durham University were first used as the experts to review all the items in the questionnaire and to look over the first draft. The draft items were changed following their suggestions to add or delete some items, and to change some of the scales used to answer the questions. The questionnaire also changed from initially asking respondents to circle numbers to tick boxes instead, due to advice given by supervisors that circling numbers may create an impression of a performance evaluation.

Next, a focus group was conducted with five female teachers from one school in the Ahmadi educational district in Kuwait, in order to discuss with teachers some problems and some particular points about the teacher evaluation system. With the focus group, this researcher intended to find out if there were any particular points that were not addressed in the first draft of the questionnaire that should have been taken into account. As a result, this

researcher confirmed that the purposes, the tools, the support of teacher evaluation system, the involvement of evaluators, etc., were all covered in the questionnaire.

The questionnaire was then prepared as a final draft and piloted with a small sample of 16 teachers from different subjects with experience ranging from 2 to 20 years. They were asked to complete the questionnaire and make notes regarding its clarity, simplicity and the time needed for its completion. This researcher found that the time taken to answer the questionnaire was between 13 and 18 minutes, which was regarded as acceptable. Moreover, it was found that, with the exception of items 4 and 6 in section two, the items were not ambiguous. The sample suggested that the terms 'formative purpose' and 'portfolio' be explained by providing some examples.

3.3.2 Interview

An interview is an instrument in which researcher and participant are involved in a conversation that is concentrated on questions related to the study. The aim of conducting an interview is to collect thoughts, perspectives, beliefs, feelings, opinions, or participants' experiences (deMarrais, 2004; Mears, 2012; Silverman, 2010). Accordingly, the interview is a flexible instrument for collecting data; researchers can address the participants' experiences, and perspectives on particular issues in order to obtain information. Through this flexibility, this researcher used the interview in order to motivate and encourage participants to provide information pertinent to the research questions, as pointed by Mack et al. (2005)

There are different types of interviews: structured, unstructured and semi-structured. This research used semi-structured interviews because the use of structured interviews with a pre-prepared list of questions allows researchers little freedom to consider anything that has not been anticipated (Cohen et al., 2007). By contrast, unstructured interviews create a more open situation and lead to more freedom for the participants, but researchers might face difficulties in collating the data because there will be more variation between interviews (Burton & Bartlett, 2009). Therefore, semi-structured interviews were used in this study to avoid the pitfalls of the other two types of interview. Thomas (2011) confirmed that semi-structured interviews provide the benefits of both unstructured and structured interviews for researchers as part of the process of collecting data. Typically, in a semi-structured interview, researchers will have a pre-prepared list of questions or topic

areas to be covered, but within this, the interviewee will be given substantial latitude to expand on their answers should they wish to do so. Questions that are not listed may be also asked based on the interviewee's responses (Bryman, 2012).

Through the semi-structured interviews that were conducted with this study's participants, this researcher intended:

- To investigate in detail the experiences, perspectives, and beliefs of the participants regarding the Kuwaiti teacher evaluation system;
- To gain their views on the alternative system proposed.

3.3.3 Focus group interviews

Patton (2002, p.385) defines a focus group interview as "an interview with a small group of people on a specific topic". A focus group is used to produce data and generate outcomes for research, by observing a group's interactions (Flick, 2009; Cohen et al., 2007).

There are several benefits to the use of focus groups that inspired this researcher to include them in this study. First, as described by Cohen et al. (2007), a focus group assists researchers in obtaining a large amount of data in a short period of time when compared to one-to-one interviews which would demand a great deal of time. Second, researchers can collect a large amount of data at a low cost by using a focus group. Third, a focus group assists researchers in gathering data that are related to attitudes, values and opinions because the participants interact more with each other than they do with the interviewer. Therefore, the increased interaction encourages more views, and thus richer data to emerge instead of researchers' own agenda. Fourth, it encourages participants to speak and to use their own words to describe, explain and introduce ideas by ensuring that all participants have opportunities to speak up and feel comfortable discussing the topic. Finally, focus groups generate diversity and difference, as noted by Flick (2009).

Consequently, a focus group was used to present the alternative teacher evaluation system to participants in order to discuss and gain their views. This was used as a means of answering the research question related to the improved usage of teacher evaluation in Kuwait.

3.4 The population and sample

As noted by Cohen et al. (2007), the quality of a piece of research can be improved through the appropriate choice of methodology and instruments, as well as by the sampling strategy that it adopts. Therefore, researchers should be aware of sampling, and make decisions about the sample at an early stage in the research plan. Before discussing the sample used for this research, the term 'population' should be explained. A population is a group of elements or cases, events, and people that conform to particular criteria or characteristics that researchers intend to study (McMillan, 1996; Mertens, 2015).

The target population for this study was teachers, head teachers in public primary schools and inspectors in the Kuwaiti MOE. According to Kuwait Central Statistical Bureau (KCBS) (2013/2014) there are 259 primary, 206 middle and 139 public high schools in Kuwait. The number of head teachers is similar to the total number of schools, as each school has one head teacher. These schools are divided amongst six educational districts (see Table 3.3).

Table 3.3: The number of schools in the educational districts

Educational Districts	Asimah	Hawalli	Ahmadi	Jahra	Farwaniya	Mubarak AlKabeer	Total
Primary schools	45	36	54	44	48	32	259
Middle schools	34	31	47	36	35	23	206
High schools	27	21	30	19	27	15	139

There are 126 primary schools for girls that have to be taught by female teachers and 133 primary schools for boys that are taught by either female or male teachers. There are 1366 male teachers and 21,376 female teachers teaching in these public primary schools across Kuwait. There are fewer male teachers than female teachers in primary education because the MOE has a policy that primary schools for boys should be taught by female teachers in the majority of schools, while only some schools for boys employ male teachers. Therefore, there are a few all-boys' schools that are taught only by male teachers in the state of Kuwait (see Table 3.4). Table 3.5 shows the number of students, classrooms, and teachers in primary schools in each educational district.

Table 3.4: The number of primary schools for boys that are taught by male and female teachers

Educational Districts	Asimah	Hawalli	Ahmadi	Jahra	Farwaniya	Mubarak AlKabeer	Total
Male	3	3	6	3	2	0	18
Female	20	16	22	19	23	16	115

Table 3.5: The number of students, classes, and teachers in the educational districts

Educational Districts	Asimah	Hawalli	Ahmadi	Jahra	Farwaniya	Mubarak AlKabeer	Total
Classrooms	811	855	1,356	1,119	1,161	639	5,941
Students	17,516	21,158	35,129	26,592	28,999	14,379	143,773
Teachers	3,229	3,249	5,162	3,964	4,475	2,663	22,742

With regard to inspectors, they work in the Departments of Inspection in the MOE. Every subject that is taught in Kuwait has a main department of inspection. The main departments are divided into six sub-departments for each of the six districts. For example, the Arabic inspection department is divided into sub-departments in the Ahmadi, Hawalli, Asimah, Jahra, Farwaniya, and Mubarak AlKabeer districts. Each sub-department has inspectors who are responsible for all types of schools in their district. This researcher was unable to obtain precise numbers of the inspectors in each sub-department or all the inspectors in the MOE, but estimates there are between 12 and 20 inspectors in each sub-department, depending on the subject taught.

3.4.1 The sample from the population

The sample is defined as the group of elements, events, or people chosen by researchers in order to collect data for the study (McMillan, 1996). The reason for sampling is to prevent researchers from having to collect data from the entire population (Cohen et al., 2007).

There are two main sampling strategies: probability and non-probability sampling. A probability sample is a randomly generated sample from the population. On the other hand, a non-probability sample is a selective sample (Cohen, et al., 2007; McMillan, 1996; Check & Schutt, 2012). This research was based on a random 'probability' sample where this researcher wishes to make generalisations from the results, since Cohen et al. (2007) indicated that probability sampling is helpful as it seeks to be representative of the population.

There are also different types of probability sampling. This researcher decided to choose cluster sampling. Cluster sampling means that researchers choose to conduct research with a naturally-occurring group of individuals randomly selected from a large population. For example, researchers choose the sample based on cities, schools, universities, and classrooms (McMillan, 1996; Mertens, 2015). Using this technique, as pointed out by Mertens (2015), researchers should choose the city randomly then test or investigate all the schools in this city or take a sample of the schools in the city in order to save time and expense.

All educational districts come under the responsibility of the MOE and the regulations for education in Kuwait (see Appendix 1); this researcher first chose three districts randomly from among the six districts: Asimah, Ahmadi and Farwaniya districts. These districts were chosen through an MOE's "Educational Research Department" application form which consisted of a list that included all names in each educational district in Kuwait; this researcher ticked three boxes of the six.

Then, three schools from each of the three districts were chosen randomly as a sample. The three schools were divided into two groups: the first group was composed of the two schools that were taught by female teachers and the other group was composed of one school that was taught by male teachers. This researcher went to the primary educational observer office in each district selected and asked for two lists of schools (one for schools taught by female teachers and two for male teachers). This researcher moved the pen between the figures in the list of schools without seeing the names of the schools, and then suddenly stopped to randomly select the school.

3.4.1.1 The sample for the questionnaire

The participants who contributed via questionnaire were all teachers. Only teachers who teach compulsory subjects were included in this research. The term 'compulsory subjects' refers to academic subjects that are taught in the national curriculum and examined during the school year. These are: Arabic, English, Science, Mathematics, Islamic Studies, Social Studies, and Computer Studies. Teachers for non-academic subjects, such as Music, Sport and Art were excluded due to the fact that these subjects are not examined and there is no specific curriculum regulating them.

The sample included nine schools within three districts. Each school received a questionnaire to be given to its teachers. The total number of teachers included in this research was 697. This researcher obtained responses from 599 teachers after discounting the incomplete questionnaires. Details regarding the responses from the teachers are found in Table 3.6.

Table 3.6: The number of teachers who participated in this research

Educational District		The Total Number of Teachers	Head of Department	Teachers with Diplomas	Teachers with Bachelor's Degrees	Teachers with Postgraduate Certificates	**Total in District**
Ahmadi	1st F/T SCHOOL	68	4	1	64	3	192
	2ND F/T SCHOOL	61	5	0	60	1	
	M/T SCHOOL	63	9	4	51	8	
Asimah	1st F/T SCHOOL	70	5	1	64	5	171
	2ND F/T SCHOOL	40	4	0	39	1	
	M/T SCHOOL	61	7	10	44	7	
Farwaniya	1st F/T SCHOOL	63	4	1	61	1	236
	2ND F/T SCHOOL	87	6	0	86	1	
	M/T SCHOOL	86	7	14	67	5	
Total:		M/T: 210 F/T: 389	51	31	537	31	599

* Note: F/T = Female teachers M/T = Male teachers

As mentioned above, the sample included all compulsory subjects in primary schools. Table 3.7 shows the number of teachers in each subject. It is possible to note that fewer Social Studies and Computer Science teachers participated in this research. This is because the number of teachers in primary schools for these two subjects is not equal to the number of teachers in other subjects because those two subjects have fewer sessions when compared to others.

Table 3.7: The number of teachers who participated, according to subject

Subject	The number of teachers
Arabic	132
English	93
Maths	98
Science	80
Islamic Studies	111
Social Studies	59
Computer Science	26
Total	599

Table 3.8 shows that 47.6% of teachers who participated in this research had less than 10 years' experience and 43.4% had between 10 and 20 years' experience. Only 9% of teachers had more than 20 years experience.

Table 3.8: The number of teachers, based on experience

Experience	Frequency	Percentages
Less than 10 years	285	47.6
Between 10 and 20	260	43.4
More than 20	54	9.0
Total	599	100%

3.4.1.2 The interview sample

The interview sample included nine head teachers, chosen from the same nine schools as the teachers. Inspectors were also chosen from the same three districts and from compulsory subjects. Once this researcher sent invitations to the inspection sub-departments in the three districts, the head inspectors nominated one inspector to participate, or one inspector volunteered to participate, in this research. However, some inspection sub-sections did not participate in this research. This researcher attempted to include inspectors from all compulsory subjects. There were 12 inspectors and two of them are the heads of inspectors in the sub-sections of the districts, as shown in Table 3.9.

This researcher chose to interview the individuals identified above because they play a significant role in the evaluation system in actively applying the criteria and enforcing the evaluation policies. An interview also allowed this researcher to obtain more in-depth data

than was obtainable from the questionnaire, due to the fact that open-ended questions could be used and adapted as necessary. Moreover, the numbers involved are such that this researcher could easily interview a group of this size.

Table 3.9: The number of inspectors who participated in this research

Inspector's subject	The number of inspectors
Arabic	2
English	2
Maths	2
Science	2
Islamic Studies	2 *
Social Studies	1
Computer Science	1 *
Total	12

*Including one district head of inspectors

3.4.1.3 The sample for focus groups

The sample was chosen from among the same schools that responded to the questionnaire. From the nine schools, five teachers were chosen for the focus groups. There were two ways to choose the teachers. First, upon completion of the questionnaire, the teachers had the opportunity to write their name down or their email or send an email to or text message this researcher if they wanted to participate in the focus groups. Second, if this researcher did not recruit enough participants through the questionnaire in one school, then this researcher asked the administrative staff or head teachers to nominate teachers from different compulsory subjects to participate in the focus groups. Then, this researcher selected different teachers for each group, depending on their experience, subjects and occupation. For example, this researcher selected two teachers with between 10 and 20 years' teaching experience and two with less than 10 years of teaching experience, taken from a range of different subjects.

3.5 Data collection procedures

This section explains the procedures used to collect the data through the questionnaire, interview, and focus groups.

3.5.1 Questionnaires

Questionnaires can be administered in several ways, for example, by post, by phone, via the Internet (Cohen et al., 2007; Tymms, 2012; Check & Schutt, 2012). Questionnaires can also be administered by handing them out to participants and collecting them later (Newby,

2014). This researcher organised the administration of the questionnaires by visiting schools and asking the head teachers or their deputies/assistants to hand out questionnaires to the teachers. This researcher gave the participants approximately five days to complete the questionnaires and then collected them from the school.

3.5.2 Interviews

This researcher conducted two sets of interviews on two different days. The first set focused on the current Kuwaiti system and the second focused on the alternative system proposed. This researcher gave the participants the freedom to choose a date and time that was convenient for them. By doing so, this researcher avoided the challenges of time and place that have been noted by various authors (Mears, 2012; Askar, Jamea, Alfarra, & Hawana, 2009). Moreover, this researcher avoided asking sensitive questions in order to avoid the risk of eliciting non-responses from participants. In addition, this researcher tried to motivate them to answer the questions more in-depth by asking for details, or through verbal and non-verbal interactions, and by avoiding leading questions.

First of all, participants were asked to sign the consent form and read the information sheet before the interview commenced. Through the consent form, participants were made to understand that the interviews would be recorded only if they agreed. Most head teachers agreed to recording their interviews; however, all the inspectors (except one) refused to have their interviews recorded. For the interviews involving the participants who refused to be recorded, this researcher took notes. Both interviews were conducted in Arabic language as the official and native language in the Kuwaiti context is Arabic.

The first interview started with an explanation of the research aims, the importance of the research, and some questions about the participant's experience, subjects taught, and so on. It then concentrated on their perspectives on the current system. The length of the first interview was between 35 and 60 minutes. After concluding the discussion on the Kuwaiti system, this researcher provided a booklet with information to the participants, which described the alternative system, and asked them to read it before the second interview (the alternative system booklet in Appendix 20 [Arabic] and Chapter Six, Section 6.3 [English]).

In the second interview, this researcher started with some explanations for the participants to describe the alternative system and its procedures. Then, this researcher listened to the participants' opinions of the alternative system, asked some questions, discussed their points of view, and presented the other views; for example, if the participant agreed with some part of the alternative system then this researcher presented the opposite view in order to let the participant defend his/her view. The length of the second interview was between 25 and 60 minutes.

3.5.3 Focus group

It was difficult to carry out the focus groups on two different days due to the teachers' heavy workloads. Therefore, this researcher conducted the focus groups in two, two-hour periods on the same day – one in the early morning and one in the early afternoon. In the first part, the aim was to build a relationship between this researcher and the teachers and to present and explain the alternative system by providing a booklet with information. Teachers also had time to read the alternative system booklet to prepare any points that were not explicit or any questions before the second part of focus group. In the second part, we discussed the issues related to this system and how it applies to the Kuwaiti situation. Focus groups were conducted in Arabic language as the official and native language of Kuwait.

All participants were also given an information sheet and a consent form to sign. The consent form included approval to record the focus group. In this research, only three focus groups were recorded as some or all of the participants in the other focus groups did not allow recording. In six of the focus groups, this researcher took notes instead.

In order to give everyone the opportunity to voice their opinions and obtain the most data from the focus groups, this researcher facilitated the discussion. The focus group was conducted in a private room to make the participants feel comfortable and confident, in order to ensure that they felt free to express their views. As noted by Cohen et al. (2007), researchers should ensure that the participants feel comfortable.

3.6 Data analysis in this study

For the current study, this researcher used the Statistical Package for the Social Sciences (SPSS) version 20 to analyse the quantitative data gathered from the questionnaire.

The questionnaire was analysed by presenting the data for each item separately in section one to contrast the actual purposes of the current system and the desired purposes of teacher evaluation. Then, section two was analysed by presenting the data for each item separately in order to show how the tools of teacher evaluation are used within the current system and to show how participants thought the tools should be used.

Other items in sections three and four of the questionnaire are divided into aggregated scales that analyse the teachers' views about the involvement of evaluators, the extent to which the current teacher evaluation system supports teachers. Factor analysis was used to confirm the validity of the questionnaire scales.

Descriptive statistics were used to present categorical and ordinal scales, to describe the distribution of the data and to display summary statistics. Furthermore, t-tests and ANOVA were used as statistical tests for separate items and scales in order to analyse the differences between gender, educational districts, and groups with different experiences and subjects.

In terms of qualitative data, this researcher started by identifying the main themes to come out of the interviews and focus groups, for example, head teachers' views about the actual purposes of teacher evaluation in the current system. This researcher read the interview transcripts several times and wrote down any impressions from the data that may be relevant to the main theme, e.g. the actual purposes of the teacher evaluation system. After that, this researcher extracted some sub-themes based on the impressions that were given under each main theme. Finally, the responses from the participants were presented as a thematic analysis divided according to main themes and sub-themes.

3.7 Validity

Fraenkel, Wallen, and Hyun (2012, p.148) define validity as "the appropriateness, correctness, meaningfulness and usefulness of the specific inferences researchers make based on the data they collect". There are many different types of validity, for example, content, construct, and face (Cohen et al., 2007; Thorndike & Thorndike-Christ, 2010; Burns, 2000; Mertens, 2015). This study adopted content validly of the questionnaire to

check such as fairly, and cover elements of the main issue that is under investigation, as explained by Cohen et al. (2007).

To achieve this, the questionnaire was discussed with supervision team at Durham University, and with nine academic staff in the School of Education at Kuwait University and the School of Basic Education (PAAET) in Kuwait as well. Two academic staff at Kuwait University had conducted studies on the teacher evaluation system in Kuwait. This researcher provided them with the aims of the study, the research questions and a copy of the questionnaire. They were asked if the questions were well presented and clear, if the questions addressed the research aims and were likely to provide answers to the research questions; their comments and feedback were requested in order to make improvements and this researcher gratefully accepted their advice and considered their suggestions.

3.8 Factor analysis

This researcher intended to create three scales to measure the teachers' perspectives on the current system according to the system's support for teaching development, learning improvement, and promotions and rewards (see 3.3.1.1**)** when designing the questionnaire. However, the point of using factor analysis was to confirm/regroup variables and constructs that involve either a few or hundreds of variables, such as the items used in the questionnaire, as emphasised by Yong and Pearce (2013).

This researcher conducted factor analysis on section four of the questionnaire that consisted of 14 items. According to Coakes and Steed (2009) the correlation matrix should exceed .3, the Kaiser-Meyer-Oklin value should be above .6, and Bartlett's Test of Sphericity is significant. This researcher found the correlation matrix revealed the presence of many coefficients of.30, the Kaiser-Meyer-Oklin value was .95, and the Bartlett's Test of Sphericity was significant with p= .0001. Therefore, factorability of the correlation matrix was assumed.

Principal components analysis revealed the presence of two components with eigenvalues exceeding 1, explaining 66.40% and 10.90% of the variance, respectively. The two factors solution explained a total of 77.31% of the variance, with Component 1 contributing to 63.47% and Component 2 contributing to 13.84% of the variance.

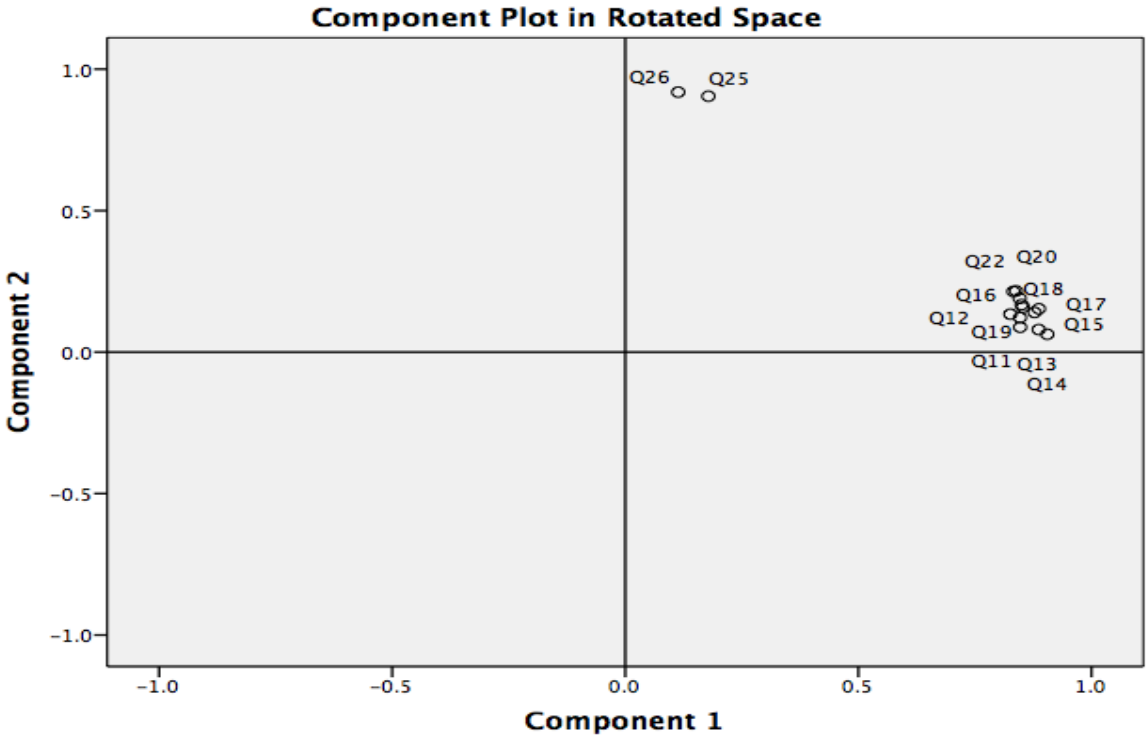
It is important to note that in Table 3.10, a score of less than .30 has been suppressed in order to make the table easy to interpret and read. The first factor consisted of 12 items in the results. The second factor consisted of two items in the results. A factor with two items is possible since "scales with more than one factor may be identified with as little as two items per factor, although these should be seen as the exception" (Raubenheimer, 2004, p.60). Furthermore, if the two items are highly correlated with each other (>.70) and uncollected with other items, the factor may be considered as reliable (Tabachinck & Fidell, 2013).

Table 3.10: Factor analysis of scales for the extent to which the current system supports teachers

Items	Component 1	Component 2
1. It has improved the deep understanding of content that you teach	.826	
2. It has assisted you with better use of pedagogies	.887	
3. It has given you a much clearer understanding of lesson planning	.905	
4. It has given you a much clearer understanding of what constitutes effective teaching	.887	
5. It has revealed the weaknesses of your performance	.850	
6. It has played a significant role in determining the strengths of your performance	.878	
7. It has affected your organisation of activities in the classroom	.847	
8. It has affected your dealing with students' discipline and behaviour problems	.853	
9. It has affected your ability to motivate students in terms of their learning	.847	
10. It has affected your ability to deal with individual differences between students	.832	
11. It has affected your continuous assessment of students' learning	.838	
12. It has affected your providing students with effective feedback	.846	
14. It has affected your rewards in terms of an annual bonus or salary increase		.904
15. It has impacted you in terms of your promotions		.919

Figure 3.1 shows that the component plot in rotated space gives one a visual representation of the loading plots in two-dimensional space. The plot shows how closely related the items are to each other and to the two components. This plot of the component loadings shows that items in factor one all load highly and positively on the first component. Items in factor two have loading near zero on the first component, but load highly on the second.

Figure 3.1: Component plot in rotated space for two factors



Consequently, this researcher amended the scales in section four of the questionnaire to include two scales. The factors are named the extent to which the current system supports the development of performance '12 items' and supports the awarding of promotions and rewards '2 items'.

3.9 Quality of scales and items

In this section, Cronbach's alpha was used to shed light on the results of reliability for scales (in sections three and four in the questionnaire) regarding their internal consistency, as suggested by Cohen et al. (2007) and Oluwatayo (2012). Analysis was also carried out to look at ceiling and flooring effects for scales (in sections three and four of the questionnaire) and items (in sections one and two of the questionnaire).

Thereafter, this researcher checked test-retest reliability with 12 teachers in two different periods of time, where aggregated scales were analysed by t-test for the mean responses between two periods of times, as explained by Cohen et al. (2007) and Oluwatayo (2012) t-test can be used. Alternatively, items were analysed by comparing the mean difference in two periods of time.

3.9.1 Quality of scales

A) Cronbach's alpha

Cohen et al. (2007, p. 506) argue that an alpha coefficient higher than .90 can be considered as very highly reliable, .80-.90 highly reliable, .70-.79 reliable, and .60-.69 marginally/minimally reliable. Following this categorisation, Table 3.11 suggests reliability for measuring the role of evaluators in providing written feedback, and discussion with teachers before and after observation. The role of head teacher, the role of inspector and the role of head of department were only .70 and above, which indicates that the scale is reliable.

Table 3.11 also shows a range of a maximum of .82 to measure rating the value of the inspector's role, which had high reliability, .78 for rating the value of the head teacher's role, and .77 for rating the value of the head of department's role that had reliability.

Table 3.11: Cronbach's alpha of scales for the involvement of evaluators

Scale		Number of Items	Cronbach's Alpha
The Role of Evaluators	Head Teacher	3	.70
	Inspector	3	.73
	Head of Department	3	.70
The Value of Evaluators' Role	Head Teacher	3	.78
	Inspector	3	.82
	Head of Department	3	.77

Table 3.12 shows Cronbach's alpha if the item was deleted for the involvement of evaluators (role and rating the value). The results show that none of the values is greater than the current alpha, except when the item *'You have had a discussion before a classroom observation'* is deleted from the component, then Cronbach's alpha increases from 0.813 to 0.892. However, this researcher decided to retain all items in the scale as this gives it more breadth.

Table 3.12: Cronbach's alpha if item deleted in scales for the involvement of evaluators

The Role of Evaluators		Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Head Teacher	You have had a discussion before classroom observation	.366	.813
	You have had a discussion after classroom observation	.604	.507
	You have received written feedback	.614	.491
Inspector	You have had a discussion before classroom observation	.366	.862
	You have had a discussion after classroom observation	.647	.539
	You have received written feedback	.694	.481
Head of Department	You have had a discussion before classroom observation	.382	.826
	You have had a discussion after classroom observation	.587	.537
	You have received written feedback	.633	.475
The Value of Evaluators' Role		**Corrected Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
Head Teacher	You have had a discussion before classroom observation	.484	.878
	You have had a discussion after classroom observation	.708	.623
	You have received written feedback	.709	.619
Inspector	You have had a discussion before classroom observation	.524	.892
	You have had a discussion after classroom observation	.737	.675
	You have received written feedback	.758	.653
Head of Department	You have had a discussion before classroom observation	.470	.866
	You have had a discussion after classroom observation	.685	.585
	You have received written feedback	.669	.597

Following this categorisation, Table 3.13 suggests reliability is high for the scale for the current system supports the development of performance (.97), and also for the scale for the current system supports the awarding of promotions and rewards (.83)

Table 3.13: Cronbach's alpha for scales for the current system supports teachers

Scale	Number of Items	Cronbach's Alpha
The system supports the development of performance	12	.97
The system supports the awarding of promotions and rewards	2	.83

Table 3.14 shows Cronbach's alpha if items are deleted for the two mentioned scales. The results show that none of the values is greater than the current alpha of the whole scale, and if an item is deleted, Cronbach's alpha does not change significantly.

Table 3.14: Cronbach's alpha if items are deleted for scales for the current system supports teachers

The system supports the development of performance	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
It has improved the deep understanding of content that you teach	.806	.969
It has assisted you with better use of pedagogies	.865	.968
It has given you a much clearer understanding of lesson planning	.881	.967
It has given you a much clearer understanding of what constitutes effective teaching	.878	.967
It has revealed the weaknesses of your performance	.838	.968
It has played a significant role in determining strengths of your performance	.865	.968
It has affected your organisation of activities in the classroom	.821	.969
It has affected your dealing with students' discipline and behaviour problems	.840	.968
It has affected your ability to motivate students in terms of their learning	.826	.969
It has affected your ability to deal with individual differences between students	.829	.969
It has affected your continuous assessment of students' learning'	.836	.938
It has affected your providing students with effective feedback	.839	.968
The system supports the awarding of promotions and rewards	**Corrected Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
It has affected your rewards in terms of an annual bonus or salary increase	.712	.
It has impacted you in terms of your promotions	.712	.

B) Ceiling and flooring effects

Figure 3.2 shows no ceiling or flooring effects for the scale regarding support of the system for performance development. For the ceiling scores pile up at 4 (the average responses) applying to over 60% of participants. By contrast, a score pile up at 2.5 (the average response) represents a downward direction.

Figure 3.2: Histogram of ceiling and flooring effect for the system supports the development of performance

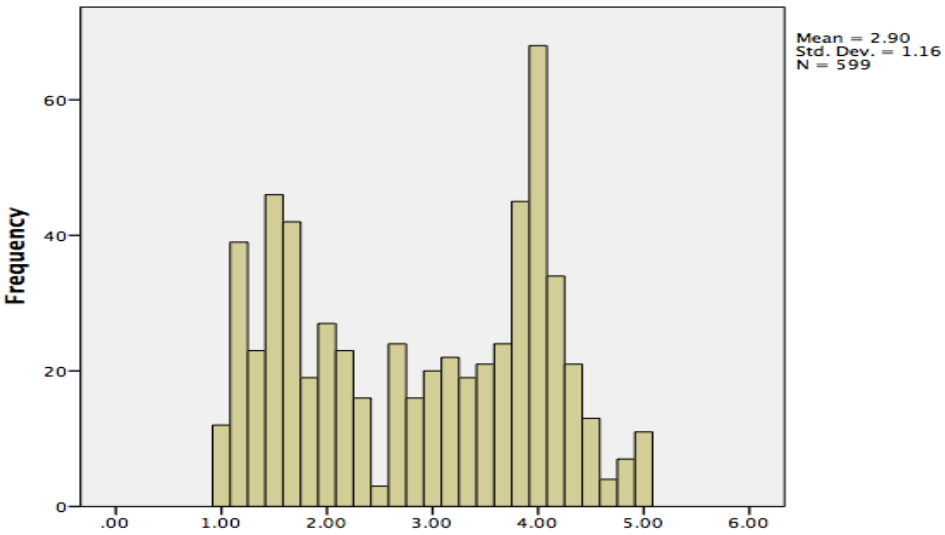
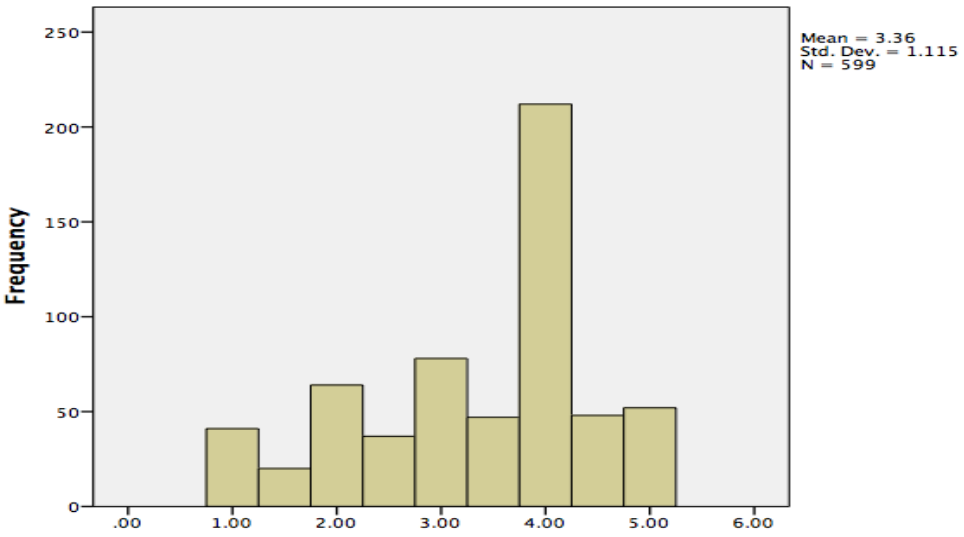


Figure 3.3 shows no ceiling or flooring effects for the scale regarding the awarding of promotions and rewards. For the ceiling scores pile up at 4 (the average response), while the scores pile up at 1.5 (the average response) as a downward direction.

Figure 3.3: Histogram of ceiling and flooring effect for the system supports the awarding of promotions and rewards



With regard to the ceiling and flooring for the involvement of evaluators in the current system (evaluators' role and rating the value of their role), the data did not show the ceiling and flooring effects (see the histograms in Appendix 23). This researcher therefore concluded that the scales had successfully avoided both floor and ceiling effects.

3.9.2 Quality of items

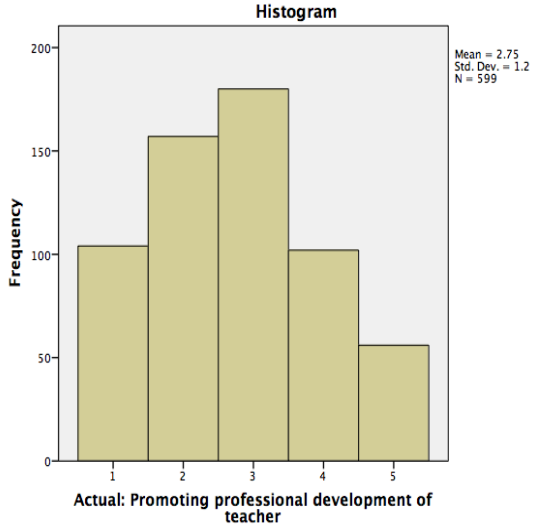
A) Ceiling and flooring effects

In sections one and two of the questionnaire, items were analysed separately. Those items were checked by looking at flooring and ceiling effects and by looking to see if all the categories were used (for a bar chart of items 1 to 9, see Appendix 21 and appendix 22). Most items show that the questions answered in those sections avoided ceiling and flooring effects. For example, in Figure 3.4, in item 1 A, it can be seen that all the categories of response have been used, where response categories 1 and 4 were chosen by 100 teachers, category 2 by around 150 teachers, category 3 by more than 150 teachers, category 5 by around 50 teachers, and modal response was 3. This pattern applies to most the questions in sections one and two, where participants' responses were contrasted, and all response options from 1 to 5 were chosen. However, the participants' responses for item 8A show

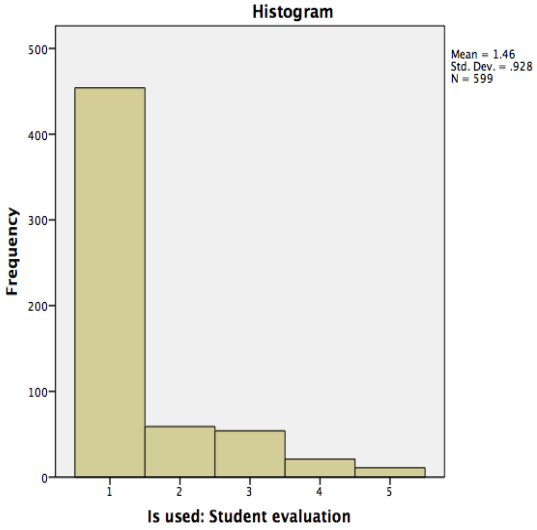
that 450 of 599 participants chose category 1 to answer this question. One point should be highlighted here regarding ceiling and flooring. If there were ceiling and flooring effects with a given item, this would not necessarily be a problem, since some of the questions were gathering facts rather than measuring attitudes. Item 8A was intended to measure the extent to which student evaluation is currently used. Most of the teachers selected never, which reflected practice rather than their views. On the other side, item 8B (see Appendix 22) asked teachers about the extent to which student evaluation should be used; teacher responses varied from 1 to 5 as this item was gathering their attitudes about using student evaluation in the evaluation of their performance.

Figure 3.4: Teachers' responses for items in sections one and two

Item 1 A



Item 8 A



3.9.3 Test-retest

Generally, researchers should consider the period of test and re-apply the test again. The period between the first and second should not be so long that responses might change or be so short that participants may remember their previous answers (Cooper & Schindler, 2001). For this test-retest, this researcher had hoped for twenty responses both times, but this was not achieved as only twelve teachers returned the second test. The results for the small sample of 12 teachers were compared between the first test and second test. This is a limitation of the test-retest mechanism: loss or lack of participants and a resulting small sample size (Creswell, 2012). A sample with less than 20 participants is too small; the sample should be as large as the researcher can obtain to provide sufficient data (Fraenkel et al., 2012). However, test-retest reliability coefficients are affected by the length of time

between the two administrations of the test (Fraenkel et al., 2012); in this case, the period between the test and retest was ten days.

3.9.3.1 Comparing mean values for two periods

With the items, Table 3.15 shows that test-retest is very good. Most of the participants responded with either the exact same category or the category above or below (difference of 1 or 2). However, with a small sample size, caution must be applied as the findings might be occurring by chance.

Table 3.15: The difference between teachers' responses in test-retest

Items		The same category	Difference of 1	Difference of 2	Difference of 3	Difference of 4	Difference of 5	Minimum	Maximum	Difference in Mean
1a	(N)%	(4) 33.3%	(2) 100%	-	-	-	-	-1.00	1.00	.33
1b	(N)%	(9) 75%	(2) 91.6%	-	(1) 100%	-	-	-1.00	3.00	.25
2a	(N)%	(7) 58.3%	(3) 83.3%	(2) 100%	-	-	-	-2.00	2.00	.25
2b	(N)%	(7) 58.3%	(4) 91.6%	-	(1) 100%	-	-	-1.00	3.00	.08
3a	(N)%	(3) 25%	(8) 91.7%	(1) 100%	-	-	-	-1.00	2.00	.50
3b	(N)%	(5) 41.7%	(6) 91.7%	-	(1) 100%	-	-	-1.00	3.00	.41
4a	(N)%	(5) 41.7%	(4) 75.1%	(3) 100%	-	-	-	-2.00	2.00	.16
4b	(N)%	(8) 66.7%	(4) 100%	-	-	-	-	-1.00	1.00	.00
5a	(N)%	(5) 41.7%	(5) 83.3%	(2) 100%	-	-	-	-2.00	2.00	.25
5b	(N)%	(5) 41.7%	(5) 83.3%	(2) 100%	-	-	-	-2.00	2.00	.25
6a	(N)%	(4) 33.3%	(5) 75%	(3) 100%	-	-	-	-2.00	2.00	.25
6b	(N)%	(2) 16.7%	(6) 66.7%	(3) 91.7%	(1) 100%	-	-	-2.00	3.00	.41
7a	(N)%	(1) 8.3%	(10) 91.6%	(1) 100%	-	-	-	-1.00	2.00	.33
7b	(N)%	(4) 33.3%	(6) 83.3%	(2) 100%	-	-	-	-2.00	2.00	.50
8a	(N)%	(6) 50%	(3) 75%	(3) 100%	-	-	-	-2.00	2.00	-.25
8b	(N)%	(5) 41.7%	(7) 100%	-	-	-	-	-1.00	1.00	.08
9a	(N)%	(1) 8.3%	(8) 74.9%	(3) 100%	-	-	-	-2.00	2.00	-.16
9b	(N)%	(3) 25%	(8) 91.7%	(1) 100%	-	-	-	-1.00	2.00	-.16

*(N)= Number of teachers

3.9.3.2 By t-test for scales

Test and re-test can be analysed by t-test for the whole test or for sections of the questionnaire (Cohen et al., 2007). Table 3.16 shows the results of the paired t-test of the mean responses for two periods of time for scales: there is no statistical difference between the teachers' responses in test and retest with regard to the involvement of evaluators, teachers' views about the current system supports the development of performance, or the awarding of promotions and rewards. The p-value is (>0.05) for all scales. However, these results are limited, as the sample was relatively small.

Table 3.16: Test and re-test results for scale by t-test

Paired Samples Test		Mean (difference between first & second)	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	DF	p-value
					Lower	Upper			
The Role of Evaluators	Head Teacher	.416	.83	.239	-.110	.944	1.73	11	.110
	Inspector	.222	.71	.206	-.232	.676	1.07	11	.305
	Head of Department	.250	1.10	.315	-.444	.944	.79	11	.445
The Value of Evaluators' Role	Head Teacher	.277	1.0	.283	-.346	.902	.97	11	.349
	Inspector	.027	.77	.222	-.462	.517	.12	11	.903
	Head of Department	.22	1.12	.323	-.490	.935	.68	11	.507
The system supports development of performance		.125	.48	.140	-.184	.434	.88	11	.394
The system supports the awarding of promotions and rewards		.416	.76	.220	-.068	.901	1.89	11	.085

3.10 Ethical considerations

Ethical issues should be taken into consideration in any research that depends on collecting data from individuals. As noted by Cohen et al. (2007), researchers should consider issues such as obtaining consent and acquiring access to the participants when conducting research, and confidentiality.

This researcher adhered to the same ethical standards as the British Educational Research Association [BERA]. BERA (2011) provides guidance on ethical considerations when carrying out research. First, researchers have a responsibility to the participants, such as by showing respect for any persons involved in the research and ensuring their fair treatment. One way in which to show respect is by gaining voluntary informed consent, having provided enough information so that participants can understand and agree to their participation, and by giving participants the right to withdraw at any time and for any or no reason, and the confidential and anonymous collection and storage of participants' data. Second, researchers have a responsibility to the sponsors of the research, which is demonstrated by conducting the research using methods that are fit for the aims of the undertaking and publication. Third, the researcher has a responsibility to the larger community of researchers, which is demonstrated by avoiding misconduct to ensure that the research is conducted at the highest standard. Fourth, researchers have a responsibility

to educational professionals, policy makers and the general public, meaning that researchers should seek to make public the results and communicate the findings and the significance of the research in a straightforward fashion using language that is appropriate for the intended audience.

This researcher submitted an application for ethical approval to the School of Education at Durham University. The application included the title of the research, the aims of the research, the significance of the research and the methods that were to be used to collect the data. After obtaining consent to carry out this research, this researcher asked the study's supervisor to write letters to conduct the research (Appendices 8 & 9). This researcher sent these letters and an email indicating that this researcher had obtained approval from the School of Education at Durham University to the Kuwait Cultural Office in London to gain a letter of permission to conduct the research in Kuwait (Appendix 10). These letters from Durham University and Kuwait Cultural Office were given to the MOE's 'Educational Research Department' in order to grant access to the schools in the three districts that were selected.

The approval letter from Educational Research Department was presented to the directors of the districts in order for them to write letters to the head teachers and inspectors to facilitate conducting this research in their districts (Appendices 11, 12 & 13). Then, this researcher obtained letters from directors of educational districts (Appendices 14 to 19), and these were presented to the head teachers of the schools chosen. Moreover, the letters were shown to inspectors and head teachers when this researcher requested their participation in this research. The letters were also shown to the teachers when they were asked to participate in the focus groups

This researcher informed the questionnaire's participants about the aims of the research by including its details in the cover page. While those who participated in interviews and focus groups were informed by providing an information sheet (Information sheet in Appendices 4 [English] and 5 [Arabic]). This researcher informed the participants that all information collected would be used for academic purposes only, with no mention being made of any type of personal information. This researcher informed them that the research would not cause them harm. In addition, this researcher informed the participants that they

were free to withdraw from participating in the research at any time. They were also informed that their views and information would not be used if they were to withdraw.

Furthermore, when asking for the consent of the participants in the interviews/focus groups (Consent form in Appendices 6 [English] and 7 [Arabic]), this researcher also asked the participants to engage in the research freely and with conviction. This researcher promised the participants that their personal information would be kept confidential, as well as any recorded information if they consented to its collection. Participants were informed that all recordings would be deleted after transcribing the data. This researcher took into account the duration of the interviews and focus group length and avoided exceeding the set time, recognising that the participants had other responsibilities to which they needed to attend.

Finally, in order to maintain anonymity of the participants in the questionnaire, this researcher did not ask for personal details from the participants, except from those who volunteered to participate in the focus groups. These participants were asked to write their email address or name at the end of the questionnaire or to send a text message or email directly to this researcher. This was used strictly for communication purposes to arrange his/her entrance in the focus groups.

Chapter Four: Teachers' Perspectives on the Current Teacher Evaluation System

4.1 Introduction

This chapter aims to analyse the data gathered in the questionnaire given to teachers asking their perspectives on the current teacher evaluation system. The chapter will begin by presenting a descriptive analysis of the data and next consider inferential statistics by looking at statistically significant differences between sub groups in background variables such as gender, experiences in teaching, subjects, and educational districts.

Before describing the data, it should be noted that not all scales were normally distributed (see Appendices 24, 25, 26, 27, & 28). However, a parametric test was applied when comparing mean values, which is described as follows:

> Naturally, since normality was assumed in the mathematical derivation of the t-test, researchers also assumed that unless the observations were normally distributed the t-test would not be a legitimate statistical option. Fortunately in recent decades subsequent research has revealed that the violation of the assumption of normality does not nullify the validity of the t-test. (Hopkins, Hopkins, & Glass, 1996, p.202)

Moreover, according to Norusis (2008, p.309), "the analysis of variance is not heavily dependent on the normality assumption". Norman (2010) stated that researchers can use parametric statistics when analysing Likert-type data with non-normal distribution.

In this research, the results of both types of tests for most pairs are the same and equivalent. For example, Tables 4.1 and 4.2 demonstrate that non-parametric and parametric test results for the actual purposes of teacher evaluation system with regard to the difference between genders show no statistical significance. The p-value is >.05 for each item.

Table 4.1: Non-parametric tests results (Mann-Whitney U) for actual purposes

Actual purposes	Grouping variable	N	Mean rank	Sum of ranks	Z	p-value
Professional development	Male	210	282.22	59267.00	– 1.90	.057
	Female	389	309.60	120433.00		
Determining performance	Male	210	288.87	60663.00	– 1.21	.227
	Female	389	306.01	119037.00		
Sanction and rewards	Male	210	304.79	64005.00	- 0.52	.603
	Female	389	297.42	115695.00		

Table 4.2: Parametric tests results (The independent samples t-test) for actual purposes

Actual purposes	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean	p-value
Professional development	Male	210	2.63	1.20	.081	.086
	Female	389	2.81	1.21	.062	
Determining performance	Male	210	3.54	1.11	.076	.172
	Female	389	3.66	1.01	.051	
Sanction and rewards	Male	210	3.81	1.16	.080	.940
	Female	389	3.81	1.06	.054	

Accordingly, parametric tests were used, as it is useful to determine the difference between variables and it is familiar to many researchers as a means to interpret and understand the data. Therefore, to test the effect of the dichotomous variables of gender, the independent samples t-test was used at 5% significance level (<.05). To test the effect variables with more than two sub-groups, such as educational districts, subjects and experience in teaching, the one-way ANOVA test was used at the 5% significance level (<.05).

4.2 Descriptive analysis

This section presents the teachers' perspectives on the purposes of teacher evaluation, the tools of teacher evaluation and the involvement of evaluators. In addition, teachers' views on the extent to which the current system supports them are presented.

4.2.1 Actual and desired purposes of the teacher evaluation system

Table 4.3 shows what teachers see as the actual purposes of the current teacher evaluation system. Two opposite patterns are revealed. The first purpose, regarding promoting the professional development of teachers, has a skewed distribution towards the lower end of the scale. The mean value for the sample is 2.75 (SD=1.20). The next two purposes, regarding determining teachers' performances and making decisions about sanctions and rewards, both have skewed distributions at the higher end. The mean values are 3.62

(SD=1.05) and 3.81 (SD=1.10), respectively. More teachers, in other words, share the view that teacher evaluation is used for determining performance and making decisions about sanctions and rewards rather than promoting professional development. Standard deviation is also higher for the professional development purpose, suggesting that there are more varied views on this as an actual purpose of teacher evaluation.

Table 4.3: Teacher's views about the actual purposes of teacher evaluation

The actual purposes of teacher evaluation		Never	Seldom	Sometimes	Often	Always	Mean	S. Deviation
Promoting professional development of teacher	Frequency	104	157	180	102	56	2.75	1.20
	%	17.4	26.2	30.1	17.0	9.3		
Determining the teacher's performance	Frequency	28	48	175	222	126	3.62	1.05
	%	4.7	8.0	29.2	37.1	21.0		
Supporting decision-makers to make decisions about teachers that are related to sanctions or rewards	Frequency	29	40	135	207	188	3.81	1.10
	%	4.8	6.7	22.5	34.5	31.4		

Interestingly, all mean values in Table 4.4, which shows teachers' desired purposes for teacher evaluation, are higher than the values observed in Table 4.3. This may be because of the different labels in the scale (i.e. importance rather than frequency), but it is clear that teachers find that all the purposes given are important. In other words, there is a contrast between desired purposes and observed in the actual purposes. Teachers find the purpose of promoting professional development of teachers as the most important (M=4.26, SD=0.86), and the purpose of supporting decision-makers to make decisions about teachers that are related to sanctions or rewards as least important (M=3.93, SD=1.15), but still important on average. This time, the purpose for making decisions about sanctions and rewards has the most variation (highest standard deviation).

Table 4.4: The desired purposes of teacher evaluation from the teachers' perspectives

The desired purposes of teacher evaluation		Unimportant at all	Unimportant	Neither important nor unimportant	Important	Very important	Mean	S. Deviation
Promoting professional development of teacher	Frequency	7	18	58	246	270	4.26	.84
	%	1.2	3.0	9.7	41.1	45.1		
Determining the teacher's performance	Frequency	20	19	60	272	228	4.12	.95
	%	3.3	3.2	10.0	45.4	38.1		
Supporting decision-makers to make decisions about teachers that are related to sanctions or rewards	Frequency	33	51	66	223	226	3.93	1.15
	%	5.5	8.5	11.0	37.2	37.7		

In spite of the different labels in the scales, paired sample t-tests were used to compare teachers' views about actual purposes and their perspectives on desired purposes of the teacher evaluation system. Table 4.5 shows, as expected, that there are statistically significant differences for all three purposes when compared to frequencies. The differences are statistically significant to the .001 level for the two first purposes, but to the .05 level for the last purpose only (sanctions and rewards).

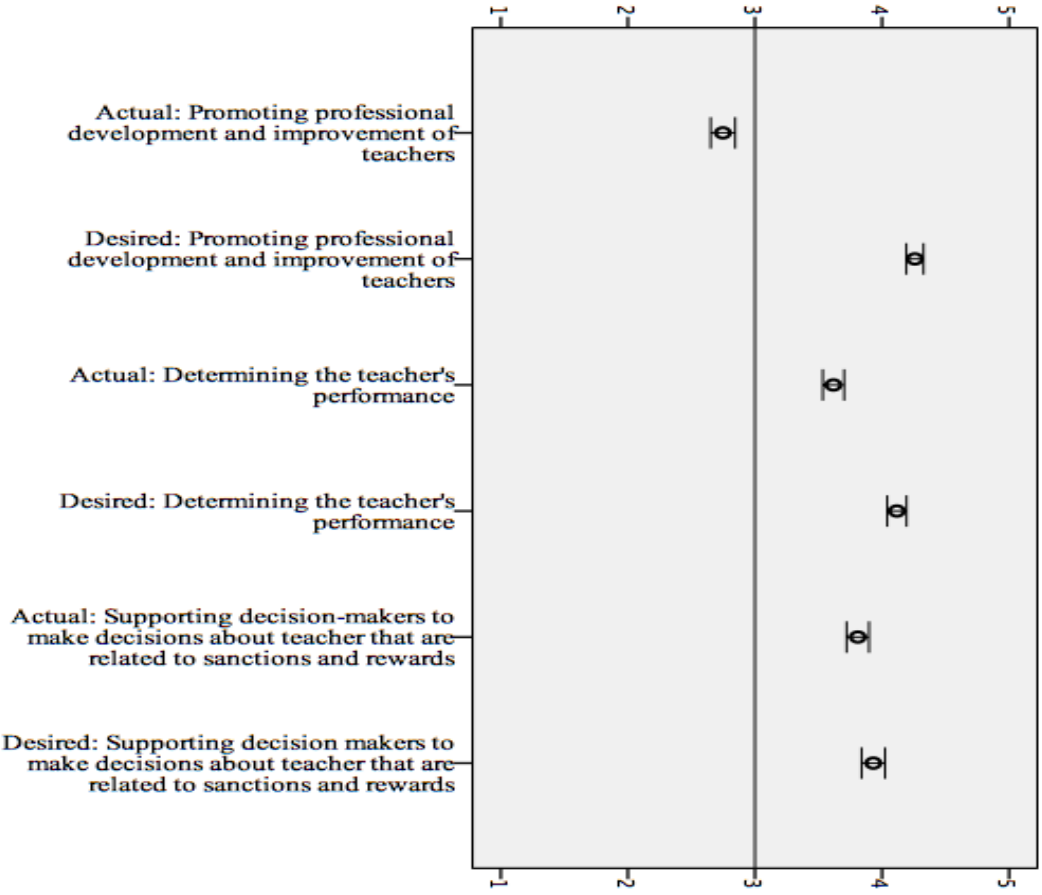
Table 4.5: The result of paired sample t-test for actual and desired purposes

Paired samples test Actual and desired purpose	Mean	Std. Deviation	Std. Error Mean	95% Confidence interval of the difference Lower	Upper	t	df	p-value
Professional development	-1.51	1.50	.061	-1.63	-1.40	-24.90	598	.000
Determining teacher performance	-.49	1.10	.045	-.58	-.41	-11.05	598	.000
Sanctions and rewards	-.12	1.35	.055	-.23	-.01	-2.20	598	.028

Figure 4.1 shows a comparison between the mean values for both the actual and desired purposes of teacher evaluation. The differences between the actual and desired purposes regarding promoting professional development are apparent. It can be seen that the other two purposes (determining performance, and making decisions about sanctions and

rewards) are similar, but not quite the same when it comes to their actual and desired purposes.

Figure 4.1: Mean and 95% CI of standard error of mean for actual and desired purposes



4.2.2 Tools of teacher evaluation that are used and should be used

Data is presented for teachers' perspectives on the different tools used in teacher evaluation. As in the previous section, teachers were first asked which tools *are* currently being used and then which tools they think *should* be used.

Table 4.6 shows the data for the tools the teachers say are used in the current system. The teachers indicate that observation is the most frequent tool used for evaluating their performance, with 84.8% of the teachers stating that this is used 'often' or 'always'. The least used tool is student evaluation, which 85.6% of teachers say is used 'never' or 'seldom'. Teachers' views are consistent with regards to the most frequently used tool being observation (SD= .85) and the least used as being student evaluation (SD= .93).

Other tools fall somewhere in between the extremes of observation and student evaluation. When it comes to student achievements and teacher portfolios, the distributions for teachers' responses are clustered from 'seldom' to 'often', and the model responses of these tools are the 'sometimes' option. When it comes to peer evaluation for formative purpose and self-evaluation, the distribution is clustered from 'never' to 'sometimes'. Here, the data show the most variation, which means that teachers' views are more divided on the frequency of those tools (high standard deviation).

Table 4.6: Teachers' views about the tools of teacher evaluation that are used.

The tools of teacher evaluation that are used		Never	Seldom	Sometimes	Often	Always	Mean	S. Deviation
Observation	Frequency	9	15	67	249	259	4.23	0.85
	%	1.5	2.5	11.2	41.6	43.2		
Student achievement	Frequency	78	141	173	126	81	2.98	1.20
	%	13.0	23.5	28.9	21.0	13.5		
Self-evaluation	Frequency	129	156	163	101	50	2.64	1.22
	%	21.5	26.0	27.2	16.9	8.3		
Peer evaluation for formative purpose	Frequency	131	175	154	81	58	2.60	1.23
	%	21.9	29.2	25.7	13.5	9.7		
Student evaluation by survey or by focus group interview	Frequency	454	59	54	21	11	1.46	0.93
	%	75.8	9.8	9.0	3.5	1.8		
Teacher portfolios	Frequency	101	134	154	121	89	2.94	1.30
	%	16.9	22.4	25.7	20.2	14.9		

It is interesting to compare what tools teachers say *are* being used in the current evaluation system as opposed to what they think should be used. Teachers give the highest priority to observation as a means of evaluating their performance and the lowest priority to student evaluation. Here, standard deviation is highest for student evaluation (SD=1.43), thus demonstrating that the range of opinions is greatest when it comes to this tool. All means are above three for observation, student achievements, student evaluation, self- and peer evaluation, and teacher portfolios (see Table 4.7).

Except for observation, which is basically the same, all means for the tools of teacher evaluation that should be used are generally higher than they are in the results on what are used. Teachers, it seems, want a broader range of evaluation tools to be used than are currently used today, and they want to participate in the evaluation (Figure 4.2). It can also be noted that there is a great discrepancy between the use of students' evaluation in the current system (M=1.46) versus the ideal system (M=3.45).

Table 4.7: The tools of teacher evaluation that should be used from the teachers' perspectives

The tools of teacher evaluation that should be used		Never	Seldom	Sometimes	Often	Always	Mean	S. Deviation
Observation	Frequency	16	26	92	185	280	4.15	1.00
	%	2.7	4.3	15.4	30.9	46.7		
Student achievement	Frequency	25	32	135	192	215	3.90	1.00
	%	4.2	5.3	22.5	32.1	35.9		
Self-evaluation	Frequency	24	42	119	224	190	3.86	1.07
	%	4.0	7.0	19.9	37.4	31.7		
Peer evaluation for formative purpose	Frequency	45	46	130	185	193	3.73	1.20
	%	7.5	7.7	21.7	30.9	32.2		
Student evaluation by survey or by focus group interview	Frequency	99	51	117	143	189	3.45	1.43
	%	16.5	8.5	19.5	23.9	31.6		
Teacher portfolios	Frequency	42	31	116	175	235	3.88	1.19
	%	7.0	5.2	19.4	29.2	39.2		

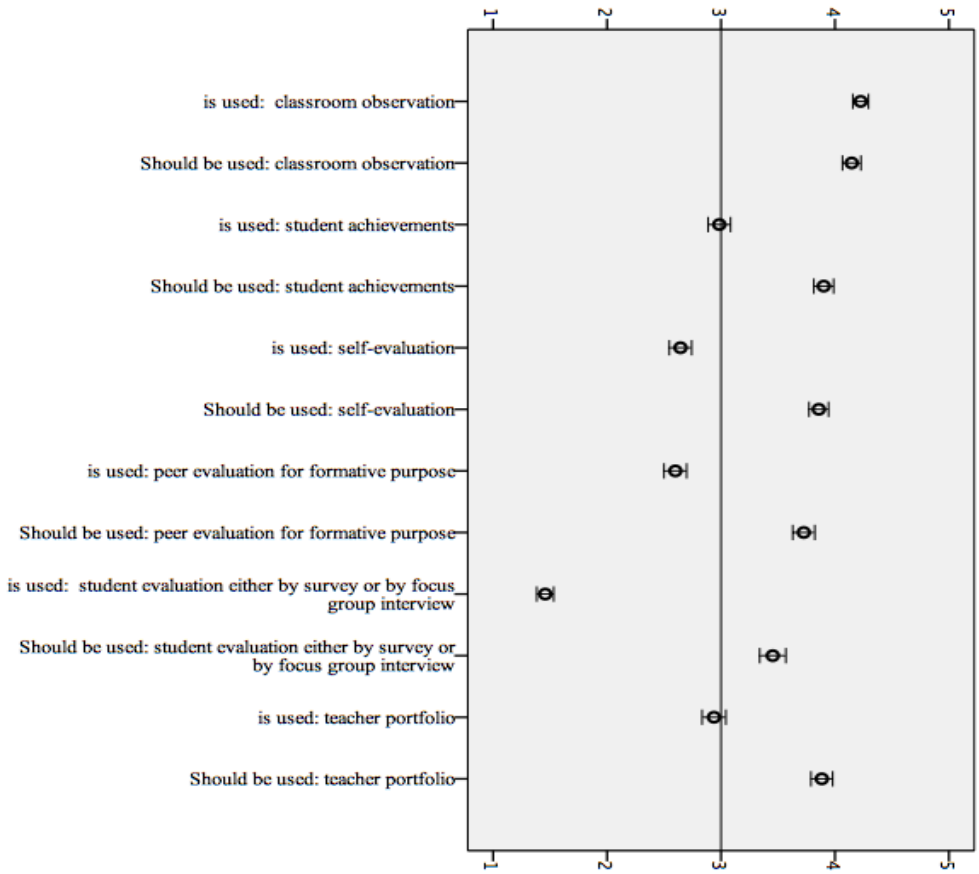
* Cited in Almutairi, Tymms, & Kind (2015, p.326-3).

Table 4.8 shows the results of a paired sample t-test between the tools that teachers indicate are used and the tools they think should be used. There is no statistically significant difference between teachers' responses regarding observation, with a p-value of .06. On the other hand, there are statistically significant differences between teachers' responses about using student achievement, self-evaluation, peer evaluation for formative purposes, student evaluation, and teacher portfolios, to a .001 level, thus supporting the view that teachers desire a greater range of evaluation tools than is used in the current system.

Table 4.8: The result of a paired t-test for tools that are used and should be used

Paired samples test Is used and Should be use	Mean	Std. Deviation	Std. Error Mean	95% Confidence interval of the difference		t	df	p-value
				Lower	Upper			
Observation	.07	1.01	.042	-.003	.16	1.88	598	.060
Student achievement	-.91	1.60	.065	-1.04	-.78	-14.14	598	.000
Self-evaluation	-1.21	1.61	.066	-1.34	-1.08	-18.38	598	.000
Peer evaluation for formative purpose	-1.12	1.55	.064	-1.25	-1.00	-17.68	598	.000
Student evaluation	-1.99	1.62	.066	-2.12	-1.86	-30.15	598	.000
Teacher portfolios	-.94	1.66	.068	-1.08	-.81	-13.91	598	.000

Figure 4.2: Mean and 95% CI of standard error of mean for what tools that are used and what should be used



4.2.3 The involvement of evaluators in teacher evaluation

Teachers were asked about the roles played by inspectors, head teachers and heads of departments. They were also required to rate the role of evaluators.

Table 4.9 shows basically the same pattern with regard to the feedback given by the teachers on all the evaluators' roles, but with different strengths. With regard to whether evaluators discuss certain points with the teachers before a classroom observation, such as evaluation criteria regarding teaching, classroom management or students' engagement, the data show that teachers' responses regarding head teachers and inspectors are skewed toward the lowest categories. Many teachers selected 'never' and 'seldom' for head teachers (56.8%) and inspectors (70.9%) whereas 42.9% of teachers selected 'never' and 'seldom', with regard to heads of departments, 25.9% of teachers selected 'sometimes'. Here, standard deviations are high for teachers' responses regarding all evaluators and reflect a more divided view on the role of evaluators in having discussions before observing teaching.

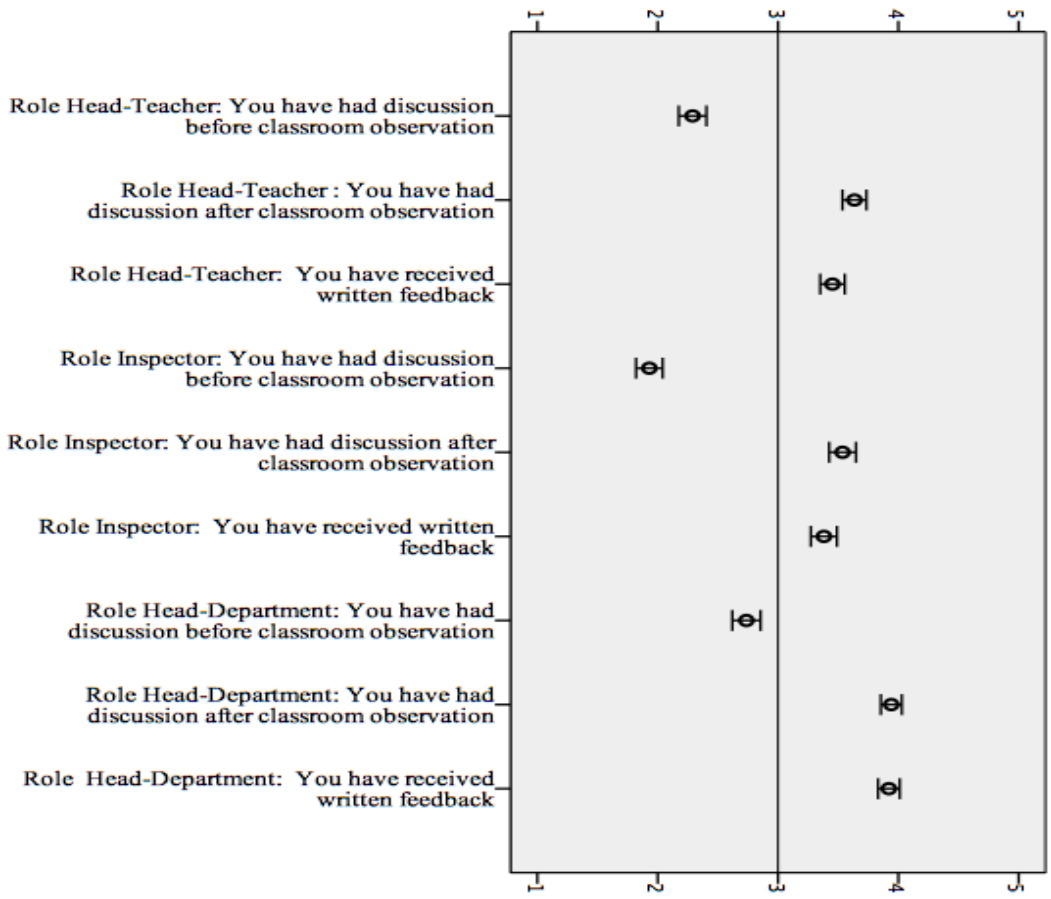
The role of evaluators in discussions with teachers *after* classroom observation and providing teachers with written feedback is in contrast to their role in discussion before observation. The distribution of teachers' responses is skewed toward the highest categories, which are 'often' and 'always'. The highest involvement of evaluators regarding these items is attributed to heads of departments (discussion after observation = 73%, written feedback = 71.1%) followed by head teachers (discussion after observation = 58.6%, written feedback = 50.6%) then inspectors (discussion after observation = 55.6%, written feedback = 50.6%). Here, opinions regarding the inspectors' role in speaking with teachers after their observations (SD=1.31) and providing them with written feedback (SD=1.30) are more divided.

Table 4.9: Teachers' views about the role of evaluators.

The Role of Evaluators								
Head Teacher		**Never**	**Seldom**	**Sometimes**	**Often**	**Always**	**Mean**	**S. Deviation**
You have had a discussion before classroom observation	Frequency	263	77	129	77	53	2.30	1.40
	%	43.9	12.9	21.5	12.9	8.8		
You have had a discussion after classroom observation	Frequency	29	83	136	173	178	3.65	1.20
	%	4.8	13.9	22.7	28.9	29.7		
You have received written feedback	Frequency	45	76	175	154	149	3.48	1.20
	%	7.5	12.7	29.2	25.7	24.9		
Inspector		**Never**	**Seldom**	**Sometimes**	**Often**	**Always**	**Mean**	**S. Deviation**
You have had a discussion before classroom observation	Frequency	344	81	75	53	46	1.96	1.32
	%	57.4	13.5	12.5	8.8	7.7		
You have had a discussion after classroom observation	Frequency	43	112	111	131	202	3.56	1.31
	%	7.2	18.7	18.5	21.9	33.7		
You have received written feedback	Frequency	51	117	128	142	161	3.41	1.30
	%	8.5	19.5	21.4	23.7	26.9		
Head of Department		**Never**	**Seldom**	**Sometimes**	**Often**	**Always**	**Mean**	**S. Deviation**
You have had a discussion before classroom observation	Frequency	156	79	142	95	76	2.74	1.40
	%	28.5	14.4	25.9	17.3	13.9		
You have had a discussion after classroom observation	Frequency	23	25	100	213	187	3.94	1.04
	%	4.2	4.6	18.2	38.9	34.1		
You have received written feedback	Frequency	26	25	107	198	192	3.92	1.10
	%	4.7	4.6	19.5	36.1	35.0		

Overall, Figure 4.3 shows that teachers perceive the evaluators as concentrating more on discussion after observation and providing written feedback. However, teachers indicate having more frequent discussions both before and after observation with the heads of departments than with the other two groups of evaluators. Teachers also received written feedback from heads of departments more frequently than from the other two evaluators.

Figure 4.3: Mean and 95% CI of standard error of mean for the role of evaluators



In Table 4.10, under the value of discussion with head teachers and inspectors before a teacher is observed, the distribution is skewed toward the lowest categories 'poor' and 'fair' for head teachers (59.6%) and for inspectors (68.1%). This result is expected due to many teachers having indicated that head teachers and inspectors have either never or seldom discussed with them before observation. The distribution of data is bimodal regarding discussion with heads of departments before observation, with 39.9% of teachers rating it as 'poor' or 'fair', and 40.7% of teachers rating it as 'very good' or 'excellent'.

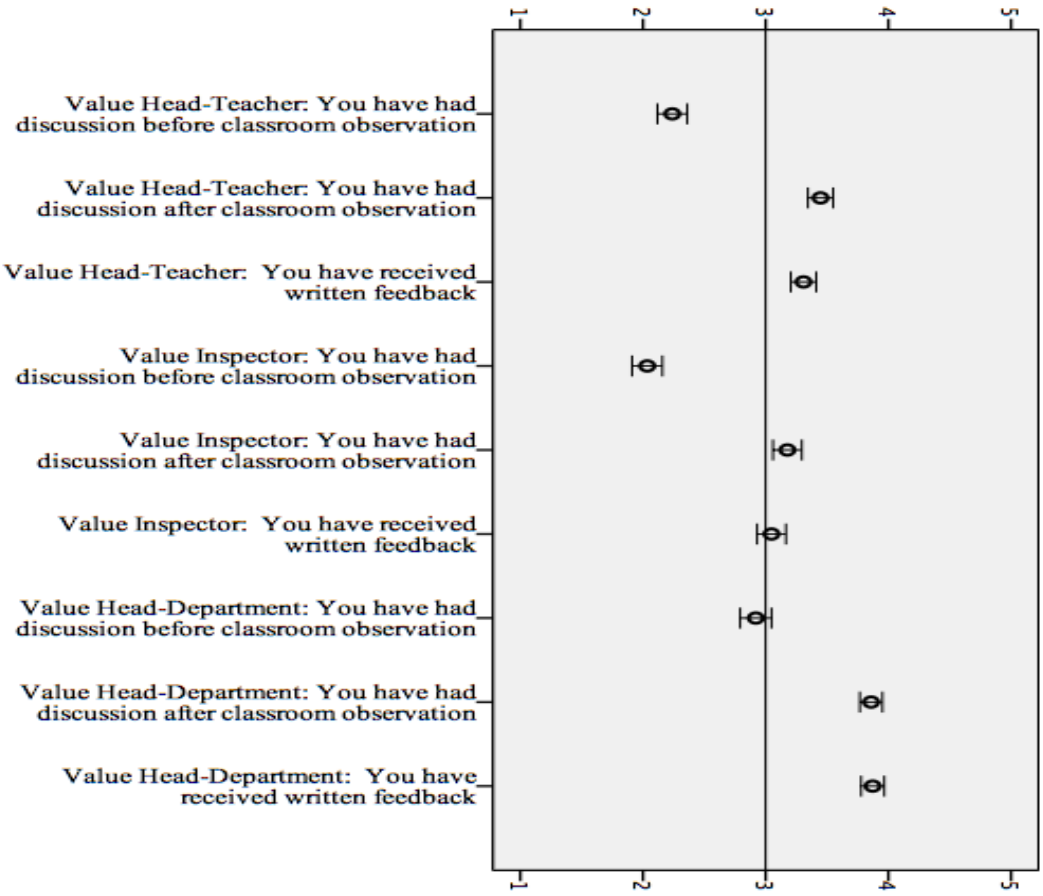
With regard to the value of discussion after observation and providing written feedback, the distributions are clustered from 'good', 'very good', to 'excellent' for both head teachers and heads of departments, but the distributions of data are bimodal for inspectors. Although the data set bimodal distribution, the data are somewhat skewed toward the highest categories, with 35.9% of teachers rating the value of inspectors' role in discussion after observation as 'poor' and 'fair', while 45% of teachers rate it as 'very good' and 'excellent'. Whereas 41.4% of teachers rate the value of written feedback from inspectors as 'poor' and 'fair', 42.4% of teachers rate it as 'very good' and 'excellent'.

Table 4.10: Teachers' rating of the value of the evaluators' role

Rating the value of evaluators' role								
Head Teacher		**Poor**	**Fair**	**Good**	**Very Good**	**Excellent**	**Mean**	**S. Deviation**
You have had a discussion before classroom observation	Frequency	295	62	105	66	71	2.26	1.45
	%	49.2	10.4	17.5	11.0	11.9		
You have had a discussion after classroom observation	Frequency	39	92	173	137	158	3.47	1.21
	%	6.5	15.4	28.9	22.9	26.4		
You have received written feedback	Frequency	58	84	187	142	128	3.33	1.23
	%	9.7	14.0	31.2	23.7	21.4		
Inspector		**Poor**	**Fair**	**Good**	**Very Good**	**Excellent**	**Mean**	**S. Deviation**
You have had a discussion before classroom observation	Frequency	357	51	62	55	74	2.06	1.50
	%	59.6	8.5	10.4	9.2	12.4		
You have had a discussion after classroom observation	Frequency	82	133	114	117	153	3.21	1.40
	%	13.7	22.2	19.0	19.5	25.5		
You have received written feedback	Frequency	92	156	97	123	131	3.08	1.40
	%	15.4	26.0	16.2	20.5	21.9		
Head of Department		**Poor**	**Fair**	**Good**	**Very Good**	**Excellent**	**Mean**	**S. Deviation**
You have had a discussion before classroom observation	Frequency	163	56	106	108	115	2.92	1.52
	%	29.7	10.2	19.3	19.7	21.0		
You have had a discussion after classroom observation	Frequency	27	33	112	194	182	3.86	1.10
	%	4.9	6.0	20.4	35.4	33.2		
You have received written feedback	Frequency	32	25	116	183	192	3.87	1.12
	%	5.8	4.6	21.2	33.4	35.0		

Overall, Figure 4.4 shows how teachers rate the value of discussion both before and after, and rate written feedback from heads of departments as more valuable than the other two groups of evaluators (head teachers and inspectors). They rate the value of the head teachers' role more highly than the inspectors' regarding discussion after observation and providing written feedback.

Figure 4.4: Mean and 95% CI of standard error of mean for rating of the value of the evaluators' role



4.2.4 Extent to which the current system supports teachers

This section is divided into two sub-sections that reflect the teachers' views about the current system in terms of the extent to which it supports development in their performance, and in awarding promotions and rewards. Here, the response categories were the five points on the Likert scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, and 5 = Strongly agree.

4.2.3.1 Extent to which the system supports the development of performance

Table 4.11 shows the frequency and mean of the teachers' responses regarding the extent to which the system supports them in developing their performance. The means of all the items in the scale are very close to the neutral point, a score of 3.00 (see Figure 4.5).

The data show a bimodal distribution for all items measuring whether the system supports them in developing their performance. This means that there is a clustering of teachers' opinions, with some teachers indicating that the current system does not support them and others indicating that it does. This can also be seen in the high standard deviations for all items (the lowest SD=1.27, the highest SD=1.42).

Teachers' responses are fairly evenly split, with 46.4% of teachers sharing the opinion that the current system has improved their deep understanding of the content that they teach and 49.3% of teachers selecting 'disagree' and 'strongly disagree'. 50.3% of teachers share the opinion that the current system has assisted them with the use of pedagogies but 44.2% of teachers 'disagree' or 'strongly disagree' with the statement. 50.4% of teachers 'agree' and 'strongly agree' that it has given them a much clearer understanding of lesson planning while 45.4% 'disagree' and 'strongly disagree' with this statement.

The current system has revealed the weaknesses in performance of 43.9% ('agree' and 'strongly agree') of the teachers while 48.3% of teachers 'disagree' and 'strongly disagree'. 48.6% of teachers 'agree' and 'strongly agree' that the current system has played a significant role in determining the strengths of their performance while 44.1% of teachers 'disagree' and 'strongly disagree'.

Regarding teachers' views on whether the current system has contributed to developing their organising activities inside the classroom, responses are equally split, with 45.9% agreeing and 45.9% disagreeing. Whether the current evaluation system gives them a clearer understanding of what constitutes effective teaching, as 46.4% of teachers disagree and 46.9% of teachers agree.

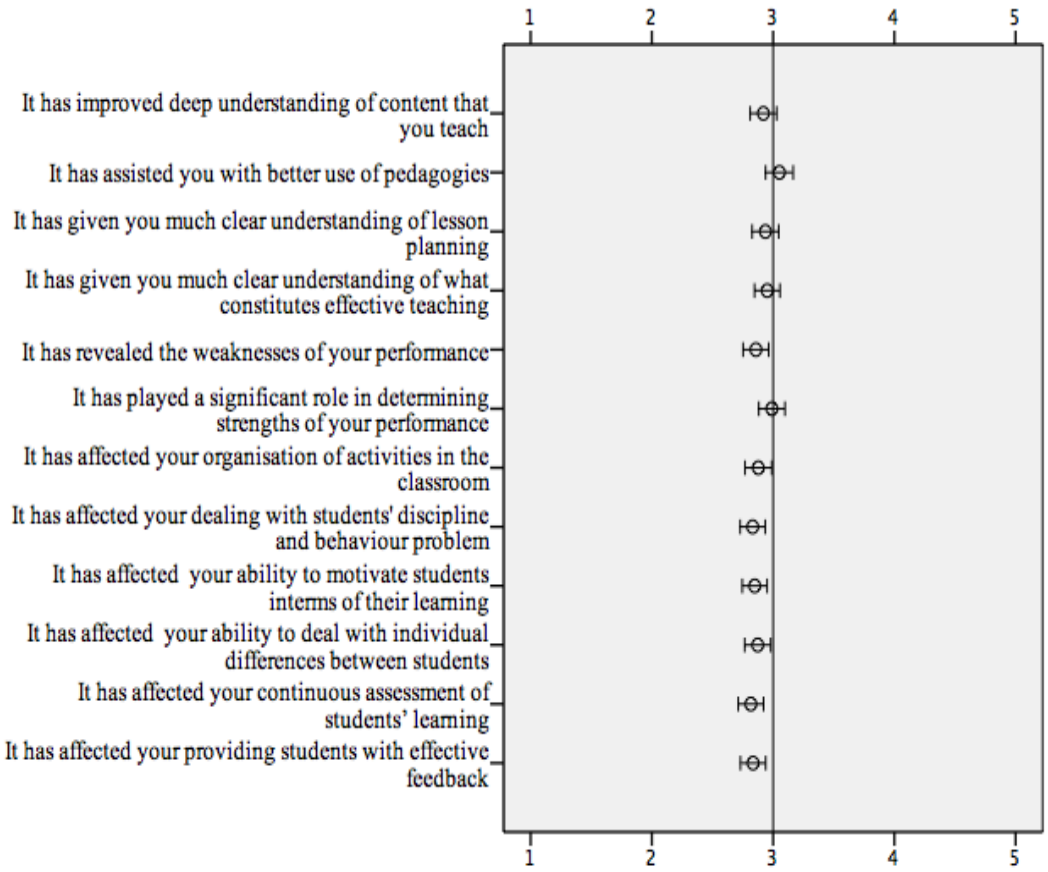
Regarding teachers' views about developing their performance with students, teachers' responses are again fairly evenly split: 48.7% of teachers disagree that the current system has affected their continuous assessment of students' learning, but 42.4% of teachers agree. 46.8% of teachers disagree on whether the current system has affected their provision of effective feedback to students by, for example, requiring teachers to monitor and record their students' progress, but 41.6% of teachers agree. Other items are placed between them. These items reflect the teachers' views on whether the current system has affected their ability to deal with students' discipline and behavioural problems, has affected their ability

to motivate students in terms of their learning, and has affected their ability to deal with individual differences between students.

Table 4.11: Teachers' views about the system supports the development of performance

The system supports the development of performance		Strongly Disagree	Disagree	Neither agree nor	Agree	Strongly Agree	Mean	S. Deviation
It has improved the deep understanding of content that you teach	Frequency	106	189	26	203	75	2.92	1.36
	%	17.7	31.6	4.3	33.9	12.5		
It has assisted you with better use of pedagogies	Frequency	109	156	33	197	104	3.08	1.42
	%	18.2	26.0	5.5	32.9	17.4		
It has given you a much clearer understanding of lesson planning	Frequency	125	147	25	245	57	2.94	1.37
	%	20.9	24.5	4.2	40.9	9.5		
It has given you a much clearer understanding of what constitutes effective teaching	Frequency	93	185	40	220	61	2.95	1.30
	%	15.5	30.9	6.7	36.7	10.2		
It has revealed the weaknesses of your performance	Frequency	113	176	47	209	54	2.86	1.32
	%	18.9	29.4	7.8	34.9	9.0		
It has played a significant role in determining the strengths of your performance	Frequency	104	160	44	220	71	2.99	1.34
	%	17.4	26.7	7.3	36.7	11.9		
It has affected your organisation of activities in the classroom	Frequency	134	141	49	214	61		
	%	22.4	23.5	8.2	35.7	10.2	2.88	1.37
It has affected your ability to deal with students' discipline and behaviour problems	Frequency	110	175	69	196	49		
	%	18.4	29.2	11.5	32.7	8.2	2.83	1.29
It has affected your ability to motivate students in terms of their learning	Frequency	107	178	53	221	40		
	%	17.9	29.7	8.8	36.9	6.7	2.85	1.27
It has affected your ability to deal with individual differences between students	Frequency	111	179	51	191	67		
	%	18.5	29.9	8.5	31.9	11.2	2.87	1.34
It has affected your continuous assessment of students' learning	Frequency	118	174	53	208	46		
	%	19.7	29.0	8.8	34.7	7.7	2.82	1.30
It has affected your providing students with effective feedback	Frequency	116	164	70	201	48		
	%	19.4	27.4	11.7	33.6	8.0	2.83	1.30

Figure 4.5: Mean and 95% CI of standard error of mean for the system supports the development of performance



4.2.3.2 Extent to which the system supports the awarding of promotions and rewards

Table 4.12 shows the frequency and mean of teachers' views about the current system when it comes to the awarding of promotions and rewards.

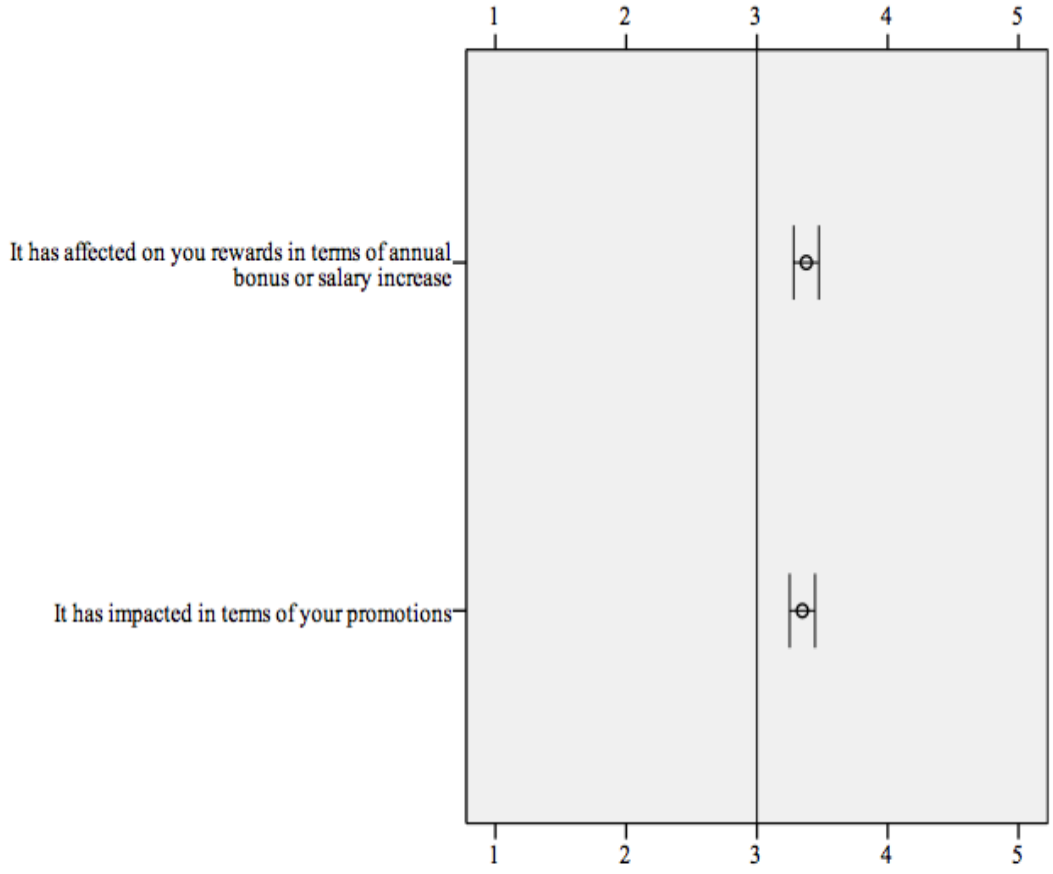
Asked if the current system is used to award bonuses and salary increases, 60.5% of teachers either agreed or strongly agreed with this item and 58.3% of teachers either agreed or strongly agreed that the current system is used to award promotions, with 27.8% of teachers either disagreeing or strongly disagreeing with this statement.

Table 4.12: Teachers' views about the system supports the awarding of promotions and rewards

The system supports the awarding of promotions and rewards		Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Mean	S. Deviation
It has affected your rewards in terms of an annual bonus or salary increase	Frequency	64	88	85	282	80	3.38	1.20
	%	10.7	14.7	14.2	47.1	13.4		
It has impacted you in terms of your promotions	Frequency	59	108	83	264	85	3.35	1.21
	%	9.8	18.0	13.9	44.1	14.2		

Overall, teachers think that the current system supports the awarding of promotions and rewards, such as annual bonuses or salary increases. The means for each item are above 3.00 and are quite similar to each other (see Figure 4.6).

Figure 4.6: Mean and 95% CI of standard error of mean for the system supports the awarding of promotions and rewards



4.2.3.3 Correlations between teachers' view about the system supports development and promotions.

Table 4.13 shows that there is a significant correlation (p=.0001) between teachers' responses regarding the system supports the development of performance, and the awarding of promotion or reward. It is apparent that these two factors affect each other positively.

Table 4.13: Spearman correlations between the system supports for development and awarding promotions

Correlation			
Spearman			Promotions
	Development	Correlation Coefficient	.298**
		Sig.	.0001
		N	599

****. Correlation is significant at the 0.01 level.**

4.3 Inferential statistics

Inferential statistics has several purposes, one of which is testing for differences between sub-groups in variables (Elst, 2013). Here, this researcher intended to discern differences between groups within the following variables: gender, educational district, extent of teaching experience, and subjects taught. Independent samples t-test and ANOVA were applied. With regard to the post hoc test to find the mean difference, both Tukey's HSD and Scheffe's tests were applied, and yielded very similar results. Therefore, Tukey's HSD post hoc test was chosen to calculate the mean difference, because it is familiar to this researcher.

4.3.1 The purposes of teacher evaluation

4.3.1.1 Gender

Table 4.14 and 4.15 show no significant difference between male and female teachers regarding the actual and desired purposes of teacher evaluation. The p-value is p>.05 for all items.

Table 4.14: The independent samples t-test of gender regarding actual purposes

Actual purposes	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean	p-value
Professional development	Male	210	2.63	1.16	.081	.086
	Female	389	2.81	1.21	.062	
Determining performance	Male	210	3.54	1.10	.076	.172
	Female	389	3.66	1.01	.051	
Sanction and rewards	Male	210	3.81	1.16	.080	.940
	Female	389	3.81	1.06	.054	

Table 4.15: The independent samples t-test of gender regarding desired purposes

Desired purposes	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean	p-value
Professional development	Male	210	4.17	.95	.066	.062
	Female	389	4.31	.76	.039	
Determining performance	Male	210	4.01	1.10	.077	.072
	Female	389	4.17	.84	.043	
Sanction and rewards	Male	210	3.84	1.23	.085	.181
	Female	389	3.98	1.10	.056	

4.3.1.2 Educational districts

Table 4.16 shows no significant difference between educational districts in terms of the actual purpose of teacher evaluation for determining performance (p>.05). However, there is a statistical difference between educational districts in terms of the actual purpose of promoting professional development, as the p-value is .025, and the actual purpose of teacher evaluation when it comes to sanctions or reward-related decision-making, with a p-value of .003.

Table 4.16: ANOVA results of educational districts regarding actual purposes

Actual purposes	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Professional development	Between Groups	10.613	2	5.307	3.719	.025
	Within Groups	850.322	596	1.427		
	Total	860.935	598			
Determining performance	Between Groups	.881	2	.440	.400	.671
	Within Groups	656.572	596	1.102		
	Total	657.452	598			
Sanction and rewards	Between Groups	8.181	2	4.090	3.414	.034
	Within Groups	714.123	596	1.198		
	Total	722.304	598			

For the post hoc analysis with regard to the actual purpose of teacher evaluation for professional development, Table 4.17 shows that teachers in Farwaniya district (M=2.84) indicate that teachers perceive teacher evaluation being used for professional development more than they are for teachers in Asimah district (M=2.54). However, the difference between the means is relatively small.

Table 4.17: Tukey's HSD post hoc test for actual purpose, educational districts

Actual purpose: professional development		N	Mean	Mean Difference	
				Asimah	Farwaniya
Educational districts	Ahmadi	192	2.82	.280	.026
	Asimah	171	2.54		305*
	Farwaniya	236	2.84		

Table 4.18 shows that teachers in Asimah (M=3.98) see the current system as being used somewhat more for the sanction and reward of teachers than do teachers in Ahmadi (M=3.68). The difference between the means is, however, relatively small

Table 4.18: Tukey's HSD post hoc test for actual purpose, educational districts

Actual purpose: sanctions and rewards		N	Mean	Mean Difference	
				Asimah	Farwaniya
Educational districts	Ahmadi	192	3.68	.300*	.120
	Asimah	171	3.98		.180
	Farwaniya	236	3.80		

The ANOVA results in Table 4.19 shows no difference between teachers in educational districts in terms of their desired purpose for determining performance, and sanctions and rewards have a p-value of >.05, but there is a statistically significant difference between educational districts in terms of the desired purpose of professional development with a p-value of <.05.

Table 4.19: ANOVA results of educational districts regarding desired purposes

Desired purposes	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Professional development	Between Groups	5.145	2	2.572	3.670	.026
	Within Groups	417.747	596	.701		
	Total	422.891	598			
Determining performance	Between Groups	2.992	2	1.496	1.674	.188
	Within Groups	532.827	596	.894		
	Total	535.820	598			
Sanction and rewards	Between Groups	5.659	2	2.830	2.150	.117
	Within Groups	784.534	596	1.316		
	Total	790.194	598			

The main difference is between educational districts with regard to the desired purpose for professional development; Table 4.20 shows that teachers in Asimah (M=4.34) believe that teacher evaluation should be used for professional development more so than teachers in Ahmadi district (M=4.13). The difference between the means is small and all means are above four.

Table 4.20: Tukey's HSD post hoc test for desired purpose, educational districts

Desired purpose: Professional development		N	Mean	Mean Difference	
				Asimah	Farwaniya
Educational districts	Ahmadi	192	4.13	.214*	.184
	Asimah	171	4.34		.030
	Farwaniya	236	4.31		

4.3.1.3 Years of teaching experience

The results of the ANOVA test in Table 4.21 show no significant difference between the different experience groups regarding the actual purposes of sanctions and rewards and determining performance, p >.05. However, there is a statistically significant difference between the different experience groups regarding the use of teacher evaluation for promoting professional development. The p-value is .0001.

Table 4.21: ANOVA results of experience groups regarding actual purposes

Actual purposes	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Professional development	Between Groups	36.740	2	18.370	13.284	.000
	Within Groups	824.195	596	1.383		
	Total	860.935	598			
Determining performance	Between Groups	.654	2	.327	.297	.743
	Within Groups	656.799	596	1.102		
	Total	657.452	598			
Sanction and rewards	Between Groups	.108	2	.054	.045	.956
	Within Groups	722.196	596	1.212		
	Total	722.304	598			

Table 4.22 shows that the main difference with regard to the actual purpose of teacher evaluation for professional development is found between teachers with less than 10 years of experience (M=2.51) and those with between 10 and 20 years of experience (M=2.90) and more than 20 years (M= 3.26). That means that the more teaching experience they have, the more they see teacher evaluation as a means of promoting professional development. This is as important as the significance between groups. The differences also

are relatively large when compared to any differences between the variables regarding actual purposes.

Table 4.22: Tukey's HSD post hoc test for actual purpose, experience groups

Actual purpose: professional development		N	Mean	Mean Difference	
				Between 10 and 20	More than 20 years
Experience groups	Less than 10 years	285	2.51	.395*	.750*
	Between 10 and 20	260	2.90		.355
	More than 20 years	54	3.26		

With regard to the desired purposes, the results in Table 4.23 show no significant difference between experience groups with regard to the desired purposes for professional development, and sanctions and rewards, with a p-value of >.05. However, there is a statistically significant difference between experience groups when it comes to teachers' perspectives on the desired purpose for determining performance, with a p-value of .037.

Table 4.23: ANOVA results of experience group regarding desired purposes

Desired purpose	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Professional development	Between Groups	1.339	2	.670	.947	.389
	Within Groups	421.552	596	.707		
	Total	422.891	598			
Determining performance	Between Groups	5.905	2	2.953	3.321	.037
	Within Groups	529.915	596	.889		
	Total	535.820	598			
Sanction and rewards	Between Groups	3.139	2	1.569	1.188	.305
	Within Groups	787.055	596	1.321		
	Total	790.194	598			

Table 4.24 shows that the main difference with regard to the desired purpose for determining performance is found between teachers with less than 10 years of experience (M=4.18) and teachers with more than 20 years of experience (M=3.81). The difference is relatively large when compared to any differences between all the variables regarding the desired purposes, however, both support this purpose. Moreover, there is no difference between the two groups 'between 10 and 20 years' and 'more than 20', which is perhaps explained by the small number of teachers among those surveyed with more than 20 years of experience.

Table 4.24: Tukey's HSD post hoc test for desired purpose, experience groups.

Desired purpose: determining performance		N	Mean	Mean Difference	
				Between 10 and 20	More than 20 years
Experience groups	Less than 10 years	285	4.18	.060	.361*
	Between 10 and 20	260	4.12		301
	More than 20 years	54	3.81		

4.3.1.4 Subjects

The results of the ANOVA tests in Tables 4.25 and 4.26 indicate that there is no significant difference between teachers in terms of subjects taught when it comes to both the actual and desired purposes of teacher evaluation, with p >.05.

Table 4.25: ANOVA results of subjects regarding actual purposes

Actual purposes	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Professional development	Between Groups	10.672	6	1.779	1.238	.285
	Within Groups	850.263	592	1.436		
	Total	860.935	598			
Determining performance	Between Groups	10.428	6	1.738	1.590	.148
	Within Groups	647.025	592	1.093		
	Total	657.452	598			
Sanction and rewards	Between Groups	14.311	6	2.385	1.994	.065
	Within Groups	707.993	592	1.196		
	Total	722.304	598			

Table 4.26: ANOVA results of subjects regarding desired purposes

Desired purposes	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Professional development	Between Groups	4.552	6	.759	1.073	.377
	Within Groups	418.340	592	.707		
	Total	422.891	598			
Determining performance	Between Groups	5.178	6	.863	.963	.450
	Within Groups	530.642	592	.896		
	Total	535.820	598			
Sanction and rewards	Between Groups	5.135	6	.856	.645	.694
	Within Groups	785.058	592	1.326		
	Total	790.194	598			

4.3.2 Tools of teacher evaluation

4.3.2.1 Gender

Table 4.27 shows the results of the independent samples t-test for gender regarding the teacher evaluation tools that are used. There is no significant difference between male and female with regard to most tools. Here, the p-value for each tool is p >.05. However, there

is a statistically significant difference between male and female perceptions with regard to observation, represented by a p-value of .004. Observation is reported as the tool of teacher evaluation that is used more often to evaluate female teachers (M=4.31) than male teachers (M=4.08). However, the difference is relatively small between genders and all means are above four.

Table 4.27: The independent samples t-test of gender regarding tools that are used

Tools of teacher evaluation that are used	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean	p-value
Observation	Male	210	4.08	1.00	.069	.004
	Female	389	4.31	.75	.038	
Student achievement	Male	210	3.01	1.30	.090	.668
	Female	389	2.97	1.20	.060	
Self-evaluation	Male	210	2.63	1.30	.089	.816
	Female	389	2.65	1.20	.060	
Peer evaluation for formative purpose	Male	210	2.62	1.30	.088	.722
	Female	389	2.59	1.22	.062	
Student evaluation	Male	210	1.49	1.00	.069	.584
	Female	389	1.44	.90	.045	
Teacher portfolios	Male	210	2.80	1.30	.088	.065
	Female	389	3.01	1.31	.066	

Similarly, Table 4.28 demonstrates that there is no significant difference between males and females with regard to all teacher evaluation tools in terms of what should be used to evaluate them. Here, each item is $p > .05$.

Table 4.28: The independent samples t-test of gender regarding tools that should be used

Tools of teacher evaluation that should be used	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean	p-value
Observation	Male	210	4.10	1.06	.073	.357
	Female	389	4.17	.97	.050	
Student achievement	Male	210	3.94	1.11	.077	.543
	Female	389	3.88	1.06	.054	
Self-evaluation	Male	210	3.85	1.10	.076	.860
	Female	389	3.86	1.05	.053	
Peer evaluation for formative purpose	Male	210	3.71	1.20	.083	.859
	Female	389	3.73	1.20	.061	
Student evaluation	Male	210	3.49	1.42	.098	.691
	Female	389	3.44	1.43	.073	
Teacher portfolios	Male	210	3.84	1.20	.083	.526
	Female	389	3.91	1.20	.060	

4.3.2.2 Educational districts

Table 4.29 shows that there is a statistical difference between educational districts in terms of the tools that are used with regard to student achievement (p=.002), self-evaluation (p=.028), and student evaluation (p=.0001). Whereas, the p-values for other tools are >.05, there is no significant difference between educational districts.

Table 4.29: ANOVA results of educational districts regarding tools that are used

Tools of teacher evaluation that are used	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Observation	Between Groups	.821	2	.410	.561	.571
	Within Groups	435.754	596	.731		
	Total	436.574	598			
Student achievement	Between Groups	19.207	2	9.603	6.477	.002
	Within Groups	883.658	596	1.483		
	Total	902.865	598			
Self-evaluation	Between Groups	10.666	2	5.333	3.585	.028
	Within Groups	886.592	596	1.488		
	Total	897.259	598			
Peer evaluation for formative purpose	Between Groups	1.713	2	.857	.559	.572
	Within Groups	914.126	596	1.534		
	Total	915.840	598			
Student evaluation	Between Groups	22.151	2	11.075	13.402	.000
	Within Groups	492.514	596	.826		
	Total	514.664	598			
Teacher portfolios	Between Groups	4.437	2	2.219	1.312	.270
	Within Groups	1008.277	596	1.692		
	Total	1012.715	598			

Table 4.30 shows the differences between educational districts when it comes to the tools that are used in teacher evaluation. Teachers in Ahmadi (M=3.21) indicated that student achievements is used in evaluating their teaching more so than the teachers in Asimah (M=2.75). There is also a difference between teachers in Ahmadi (M=2.77) and in Asimah (M=2.44) district when it comes to using self-evaluation. Moreover, teachers in Ahmadi (M=1.73) perceive student evaluation to be used more often than do other teachers in other educational districts (Asimah: M=1.29; Farwaniya: M=1.36).

Table 4.30: Tukey's HSD post hoc test for tools that are used, educational districts

Tools: are used	Educational districts	N	Mean	Mean Difference	
				Asimah	Farwaniya
Student achievement	Ahmadi	192	3.21	459*	.247
	Asimah	171	2.75		.212
	Farwaniya	236	2.97		
Self-evaluation	Ahmadi	192	2.77	.327*	.071
	Asimah	171	2.44		.256
	Farwaniya	236	2.69		
Student evaluation	Ahmadi	192	1.73	.448*	.378*
	Asimah	171	1.29		.069
	Farwaniya	236	1.36		

Table 4.31 indicates that there is a statistical difference between educational districts with regard to the extent to which the listed tools should be used in teacher evaluation. Here, each item is $p < .05$.

Table 4.31: ANOVA results of educational districts regarding tools that should be used

Tools of teacher evaluation that should be used	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Observation	Between Groups	10.541	2	5.271	5.266	.005
	Within Groups	596.531	596	1.001		
	Total	607.072	598			
Student achievement	Between Groups	20.313	2	10.156	8.943	.000
	Within Groups	676.876	596	1.136		
	Total	697.189	598			
Self-evaluation	Between Groups	24.986	2	12.493	11.351	.000
	Within Groups	655.953	596	1.101		
	Total	680.938	598			
Peer evaluation for formative purpose	Between Groups	17.051	2	8.525	5.977	.003
	Within Groups	850.048	596	1.426		
	Total	867.098	598			
Student evaluation	Between Groups	34.550	2	17.275	8.667	.000
	Within Groups	1187.937	596	1.993		
	Total	1222.487	598			
Teacher portfolios	Between Groups	13.276	2	6.638	4.757	.009
	Within Groups	831.775	596	1.396		
	Total	845.052	598			

In Table 4.32, with regard to the tools that should be used in teacher evaluation, a difference is found between Farwaniya district and the other educational districts, Ahmadi and Asimah. In other words, teachers in Farwaniya are more supportive of using a range of tools in their evaluation than teachers in the other districts. The biggest difference between educational districts is in the extent to which teachers think self-evaluation and student achievement should be used. However, there is no difference between the teachers' perspectives in Farwaniya and in Asimah regarding the use of peer evaluation, student

evaluation and teacher portfolios. It should be noted that the mean average for all tools is above three or four.

Table 4.32: Tukey's HSD post hoc test for tools that should be used, educational districts

Tools: should be used	Grouping variable	N	Mean	Mean Difference	
				Asimah	Farwaniya
Observation	Ahmadi	192	4.02	0.55	.294*
	Asimah	171	4.07		239*
	Farwaniya	236	4.31		
Student achievement	Ahmadi	192	3.68	.172	432*
	Asimah	171	3.85		261*
	Farwaniya	236	4.11		
Self-evaluation	Ahmadi	192	3.65	.109	.460*
	Asimah	171	3.75		.352*
	Farwaniya	236	4.11		
Peer evaluation for formative purpose	Ahmadi	192	3.52	.186	.400*
	Asimah	171	3.70		.213
	Farwaniya	236	3.92		
Student evaluation	Ahmadi	192	3.14	.310	.571*
	Asimah	171	3.45		.262
	Farwaniya	236	3.71		
Teacher portfolios	Ahmadi	192	3.70	.168	.353*
	Asimah	171	3.87		.185
	Farwaniya	236	4.05		

4.3.2.3 Years of teaching experience

Table 4.33 shows that there is a statistically significant difference between the groups of teachers in terms of the experience groups with regard to their views on the majority tools that are used in their evaluation. Here, students' achievements, self-evaluation, peer evaluation, and students' evaluation are $p < .05$, whereas there is no significant difference between teachers in terms of experience groups when it comes to the use of teachers' portfolios and observations as tools ($p > .05$). There are more differences between experience groups regarding tools that are used than are found for the other variables.

Table 4.33: ANOVA Results of experience group regarding tools that are used

Tools of teacher evaluation that are used	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Observation	Between Groups	4.170	2	2.085	2.874	.057
	Within Groups	432.404	596	.726		
	Total	436.574	598			
Student achievement	Between Groups	34.188	2	17.094	11.728	.000
	Within Groups	868.677	596	1.458		
	Total	902.865	598			
Self-evaluation	Between Groups	19.701	2	9.851	6.690	.001
	Within Groups	877.558	596	1.472		
	Total	897.259	598			
Peer evaluation for formative purpose	Between Groups	17.466	2	8.733	5.794	.003
	Within Groups	898.374	596	1.507		
	Total	915.840	598			
Student evaluation	Between Groups	7.691	2	3.846	4.521	.011
	Within Groups	506.973	596	.851		
	Total	514.664	598			
Teacher portfolios	Between Groups	8.615	2	4.308	2.557	.078
	Within Groups	1004.099	596	1.685		
	Total	1012.715	598			

With regards to using student achievement, self-evaluation and peer evaluation as tools, Table 4.34 shows differences between teachers with less than 10 years of experience and teachers with between 10 and 20 and with more than 20 years of experience. Therefore, it appears that teachers with more experience participate in teacher evaluation through self- and peer evaluation more often than those with less experience. They also seem to think that students participate in evaluating them and that student achievement is used as an indicator of their performance more so than less experienced teachers.

Table 4.34: Tukey's HSD post hoc test for tools that are used, experience groups

Tools: are used	Grouping variable	N	Mean	Mean Difference	
				Between 10 to 20	More 20 years
Student achievement	Less than 10 years	285	2.74	.452*	.578*
	Between 10 and 20	260	3.19		.126
	More than 20 years	54	3.31		
Self-evaluation	Less than 10 years	285	2.46	.325*	.485*
	Between 10 and 20	260	2.78		.160
	More than 20 years	54	2.94		
Peer evaluation for formative purpose	Less than 10 years	285	2.43	.295*	.479*
	Between 10 and 20	260	2.72		.184
	More than 20 years	54	2.91		
Student evaluation	Less than 10 years	285	1.37	.213*	.076
	Between 10 and 20	260	1.58		.288
	More than 20 years	54	1.30		

Table 4.35 shows no significant difference between teachers in terms of their experience and the extent to which they agree that the various tools should be used in teacher evaluation (p >.05) except for two items. The exceptions are with regard to peer evaluation (p=.012), and teacher portfolios (p=.023).

Table 4.35: ANOVA results of experience groups regarding tools that should be used

Tools of teacher evaluation that should be used	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Observation	Between Groups	1.201	2	.601	.591	.554
	Within Groups	605.871	596	1.017		
	Total	607.072	598			
Student achievement	Between Groups	2.241	2	1.121	.961	.383
	Within Groups	694.947	596	1.166		
	Total	697.189	598			
Self-evaluation	Between Groups	2.926	2	1.463	1.286	.277
	Within Groups	678.013	596	1.138		
	Total	680.938	598			
Peer evaluation for formative purpose	Between Groups	12.664	2	6.332	4.417	.012
	Within Groups	854.435	596	1.434		
	Total	867.098	598			
Student evaluation	Within Groups	6.816	2	3.408	1.671	
	Between Groups	1215.672	596	2.040		.189
	Total	1222.487	598			
Teacher portfolios	Within Groups	10.675	2	5.337	3.812	
	Total	834.377	596	1.400		.023
	Between Groups	845.052	598			

Table 4.36 shows that differences exist between teachers with less than 10 years of experience and teachers with more than 20 years of experience when it comes to peer evaluation and teacher portfolios. Here, teachers with less than 10 years of experience (M=3.86) support using peer evaluation more than teachers with more than 20 years of experience do (M=3.41). Teachers with less than 10 years of experience (M=4.01) support using teacher portfolios more than teachers with more than 20 years of experience do (M=3.61). It should be noted that the mean average for all tools is above three or four.

Table 4.36: Tukey's HSD post hoc test for tools that should be used, experience groups

Tool: should be used	Grouping variable	N	Mean	Mean Difference	
				Between 10 to 20	More 20 years
Peer evaluation for formative purpose	Less than 10 years	285	3.86	.221	.456*
	Between 10 and 20	260	3.64		.235
	More than 20 years	54	3.41		
Teacher portfolios	Less than 10 years	285	4.01	.214	.403*
	Between 10 and 20	260	3.80		.189
	More than 20 years	54	3.61		

4.3.2.4 Subjects

Table 4.37 shows no significant difference between teachers in terms of subjects regarding tools that are used in their evaluation. The p-value for each tool is >0.05, with the exception of two items. By contrast, there are statistical differences between subjects when it comes to the degree to which self-evaluation and portfolios are used as tools in teacher evaluation. For this, p <.05.

Table 4.37: ANOVA results of subjects regarding tools that are used

Tools of teacher evaluation that are used	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Observation	Between Groups	4.464	6	.744	1.019	.412
	Within Groups	432.110	592	.730		
	Total	436.574	598			
Student achievement	Between Groups	5.515	6	.919	.606	.725
	Within Groups	897.349	592	1.516		
	Total	902.865	598			
Self-evaluation	Between Groups	30.419	6	5.070	3.462	.002
	Within Groups	866.839	592	1.464		
	Total	897.259	598			
Peer evaluation for formative purpose	Between Groups	8.762	6	1.460	.953	.456
	Within Groups	907.078	592	1.532		
	Total	915.840	598			
Student evaluation	Between Groups	2.217	6	.370	.427	.861
	Within Groups	512.447	592	.866		
	Total	514.664	598			
Teacher portfolios	Between Groups	25.570	6	4.262	2.556	.019
	Within Groups	987.144	592	1.667		
	Total	1012.715	598			

The post hoc analyses presented in Table 4.38 reveal that the main differences are found between English teachers (M=2.24) and Islamic Studies teachers (M=2.78), and Social Studies teachers (M=2.90) and Computer Science teachers (M=3.19), with regard to using self-evaluation in their evaluation. English teachers rate their participation in their evaluation via self-evaluation lower than the others teachers.

Meanwhile, the difference between Islamic Studies teachers (M=2.69) and Computer Science teachers (M=3.58) with regard to teacher portfolios reveals that Computer Science teachers think that teacher portfolios are used as an aspect of their evaluation more so than Islamic Studies teachers do.

Table 4.38: Tukey's HSD post hoc test for tools that are used, subjects

Tools: are used	Grouping variable	N	Mean	Mean Difference					
				English	Islamic	Science	Math	Social	Computer
Self evaluation	Arabic	132	2.57	.332	.216	.144	.054	.330	.624
	English	93	2.24		.547*	.476	.386	.622*	.956*
	Islamic	111	2.78			.071	.161	.115	.409
	Science	80	2.71				.090	.186	.480
	Math	98	2.62					.276	.570
	Social	59	2.90						.294
	Computer	26	3.19						
Teacher portfolios	Arabic	132	3.03	.192	.337	.093	.194	.207	.547
	English	93	2.84		.145	.099	.002	.399	.738
	Islamic	111	2.69			.244	.143	.544	.883*
	Science	80	2.94				.101	.300	.639
	Math	98	2.84					.401	.740
	Social	59	3.24						.340
	Computer	26	3.58						

Table 4.39 shows that there are statistical differences between teacher in terms of subjects regarding the extent to which each of the tools should be used in teacher evaluation. Here, each tool is either $p < .01$ or $p < .05$. The differences within this variable regarding the tools that should be used are greater when compared to other variables and equal when compared to the educational districts.

Table 4.39: ANOVA results of subjects regarding tools that should be used

Tools of teacher evaluation that should be used	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Observation	Between Groups	15.168	6	2.528	2.528	.020
	Within Groups	591.904	592	1.000		
	Total	607.072	598			
Student achievements	Between Groups	19.852	6	3.309	2.892	.009
	Within Groups	677.337	592	1.144		
	Total	697.189	598			
Self-evaluation	Between Groups	14.408	6	2.401	2.133	.048
	Within Groups	666.530	592	1.126		
	Total	680.938	598			
Peer evaluation for formative purpose	Between Groups	22.548	6	3.758	2.634	.016
	Within Groups	844.551	592	1.427		
	Total	867.098	598			
Student evaluation	Between Groups	34.785	6	5.798	2.890	.009
	Within Groups	1187.702	592	2.006		
	Total	1222.487	598			
Teacher portfolios	Between Groups	24.097	6	4.016	2.896	.009
	Within Groups	820.954	592	1.387		
	Total	845.052	598			

Table 4.40 shows the differences between the opinions of Arabic teachers and others teachers regarding the extent to which tools should be used in their evaluation. Arabic teachers generally support the use of observation, student achievement, self-evaluation, and peer evaluation more so than the other subject teachers do.

The exceptions are with regard to teacher portfolio between Math teachers (M=3.65) and both Arabic teachers (M=4.04) and Islamic Studies teachers (M=4.08). Moreover, it is revealed that Islamic Studies teachers support using student evaluation (M=3.57) more so than their colleagues who teach Computer Science (M=2.81).

Table 4.40: Tukey's HSD post hoc test for tools that should be used, subjects

Tools: should be used	Grouping variable	N	Mean	Mean Difference					
				English	Islamic	Science	Math	Social	Computer
Observation	Arabic	132	4.37	.448*	.120	.368	.226	.055	.164
	English	93	3.92		.273	.025	.167	.393*	.229
	Islamic	111	4.20			.248	.106	.175	.044
	Science	80	3.95				.142	.423	.204
	Math	98	4.09					.281	.62
	Social	59	4.32						.219
	Computer	26	4.15						
Student achievements	Arabic	132	4.17	.210	.203	.442	.514*	.319	.474
	English	93	3.96		.007	.232	.304	.110	.265
	Islamic	111	3.96			.239	.311	.117	.272
	Science	80	3.73				.072	.122	.033
	Math	98	3.65					.194	.039
	Social	59	3.85						.155
	Computer	26	3.69						
Self-evaluation	Arabic	132	4.06	.136	.115	.411*	.326	.417*	.253
	English	93	3.92		.021	.275	.190	.281	.117
	Islamic	111	3.95			.296	.211	.302	.138
	Science	80	3.65				.085	.006	.158
	Math	98	3.73					.091	.073
	Social	59	3.64						.164
	Computer	26	3.81						
Peer evaluation for formative purpose	Arabic	132	3.91	.038	.071	.309	.470*	.180	.640
	English	93	3.87		.033	.271	.432	.142	.602
	Islamic	111	3.84			.238	.399	.109	.569
	Science	80	3.60				.161	.129	.311
	Math	98	3.44					.290	.170
	Social	59	3.73						.460
	Computer	26	3.27						
Student evaluation	Arabic	132	3.59	.032	.157	.403	.366	.133	.783
	English	93	3.56		.189	.372	.335	.102	.751
	Islamic	111	3.75			.560	.523	.290	.940*
	Science	80	3.19				.037	.270	.380
	Math	98	3.22					.233	.417
	Social	59	3.46						.650
	Computer	26	2.81						
Teacher portfolios	Arabic	132	4.04	.188	.043	.375	.477*	.047	.192
	English	93	3.85		.232	.187	.288	.235	.003
	Islamic	111	4.08			.419	.520*	.004	.235
	Science	80	3.66				.101	.422	.184
	Math	98	3.56					.524	.285
	Social	59	4.08						.239
	Computer	26	3.85						

4.3.3The involvement of evaluators in teacher evaluation

4.3.3.1 Gender

The results of the independent sample t-tests are shown in Table 4.41. There is no significant difference between female and male teachers in terms of the role of evaluators and rating the value of their role. The p-value for each is >.05.

Table 4.41: The independent samples t-test of gender for the involvement of evaluators.

Involvement of evaluators	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean	p-value
Role of head teachers	Male	210	3.10	1.06	.073	
	Female	389	3.16	.95	.048	.521
Value of head teachers	Male	210	3.01	1.22	.084	
	Female	389	3.02	1.01	.051	.922
Role of inspectors	Male	210	3.02	1.15	.079	
	Female	389	2.95	1.00	.050	.459
Value of inspectors	Male	210	2.81	1.26	.087	
	Female	389	2.76	1.18	.060	.589
Role of heads of departments	Male	187	3.53	1.01	.074	
	Female	361	3.53	.88	.046	.993
Value of heads of departments	Male	187	3.46	1.10	.079	
	Female	361	3.59	1.00	.052	.184

4.3.3.2 Educational districts

The results of the ANOVA test in Table 4.42 indicate no statistically significant difference between teachers in educational districts with regard to both the roles and the rating of the value of these roles for heads of departments and inspectors (>.05).

However, there is a statistical difference between teachers in educational districts with regard to the role of head teachers and rating the value of the head teachers' role. The p-value is p=.0001 for the role of head teachers, and the p-value is p=.023 for rating the value of the head teachers' role.

Table 4.42: ANOVA results of educational districts for the involvement of evaluators

Involvement of evaluators	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Role of head teachers	Between Groups	19.089	2	9.545	9.975	.000
	Within Groups	570.277	596	.957		
	Total	589.366	598			
Value of head teachers	Between Groups	8.975	2	4.487	3.815	.023
	Within Groups	701.105	596	1.176		
	Total	710.079	598			
Role of inspectors	Between Groups	6.197	2	3.098	2.775	.063
	Within Groups	665.476	596	1.117		
	Total	671.673	598			
Value of inspectors	Between Groups	3.693	2	1.847	1.249	.288
	Within Groups	881.170	596	1.478		
	Total	884.864	598			
Role of heads of departments	Between Groups	.780	2	.390	.477	.639
	Within Groups	474.940	545	.871		
	Total	475.720	547			
Value of heads of departments	Between Groups	2.705	2	1.352	1.263	.284
	Within Groups	583.787	545	1.071		
	Total	586.492	547			

Table 4.43 shows that teachers in Ahmadi (M=3.40) recognise the participation of head teachers in their evaluation more so than teachers in Asimah (M=3.02) and teachers in the Farwaniya (M=3.01) districts. Thereby, Ahmadi teachers rate the value of the head teachers' role as more valuable than do the teachers in other educational districts.

Table 4.43: Tukey's HSD post hoc test for the involvement of evaluators, educational districts

Involvement of evaluators	Grouping variable	N	Mean	Mean Difference	
				Asimah	Farwaniya
Role of head teacher	Ahmadi	192	3.40	.371*	.389*
	Asimah	171	3.02		.017
	Farwaniya	236	3.01		
Value of head teachers' role	Ahmadi	192	3.13	.304*	.079
	Asimah	171	2.83		.225
	Farwaniya	236	3.05		

4.3.3.3 Years of teaching experience

The results of the ANOVA test in Table 4.44 show that there is no statistical difference between the experience groups in terms of either the role or rating of the value of head teachers and heads of departments, with p-values of >.05 for each. However, there is a statistical difference between the experience groups with regard to the role of inspectors (p=.008) and rating the value of the inspectors' role (p=.001).

Table 4.44: ANOVA results of experience groups for the involvement of evaluators

Involvement of evaluators	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Role of head teachers	Between Groups	5.276	2	2.638	2.692	.069
	Within Groups	584.090	596	.980		
	Total	589.366	598			
Value of head teachers	Between Groups	5.218	2	2.609	2.206	.111
	Within Groups	704.861	596	1.183		
	Total	710.079	598			
Role of inspectors	Between Groups	10.884	2	5.442	4.909	.008
	Within Groups	660.789	596	1.109		
	Total	671.673	598			
Value of inspectors	Between Groups	20.756	2	10.378	7.158	.001
	Within Groups	864.108	596	1.450		
	Total	884.864	598			
Role of heads of departments	Between Groups	.834	2	.417	.478	.620
	Within Groups	474.886	545	.871		
	Total	475.720	547			
Value of heads of departments	Between Groups	.346	2	.173	.161	.851
	Within Groups	586.146	545	1.075		
	Total	586.492	547			

Table 4.45 shows that teachers with more experience are more likely than teachers with less experience to perceive inspectors as being more involved in teacher evaluation through their discussions with them both before and after observation and by providing written feedback. Therefore, teachers with more experience rate the value of the inspectors' role more highly than it is rated by less experienced teachers.

Table 4.45: Tukey's HSD post hoc test for the involvement of evaluators, experience groups

Involvement of evaluators	Grouping variable	N	Mean	Mean Difference	
				Between 10 and 20	More than 20 years
Role of inspector	Less than 10 years	285	2.85	.178	.450*
	Between 10 and 20	260	3.03		.271
	More than 20 years	54	3.30		
Value of inspector	Less than 10 years	285	2.59	.332*	.499*
	Between 10 and 20	260	2.92		.166
	More than 20 years	54	3.09		

4.3.3.3 Subjects

Table 4.46 indicates that there is a significant different between subjects with regard to the role of head teachers and heads of departments and the value of their role. Each scale is p >.05. By contrast, there are no significant differences between subjects with regard to the role of inspectors and the value of the inspectors' role (p <.05).

Table 4.46: ANOVA results of subjects for the involvement of evaluators.

Involvement of evaluators	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Role of head teachers	Between Groups	15.931	6	2.655	2.741	.012
	Within Groups	573.435	592	.969		
	Total	589.366	598			
Value of head teachers	Between Groups	28.451	6	4.742	4.118	.000
	Within Groups	681.628	592	1.151		
	Total	710.079	598			
Role of inspectors	Between Groups	10.993	6	1.832	1.642	.133
	Within Groups	660.680	592	1.116		
	Total	671.673	598			
Value of inspectors	Between Groups	12.075	6	2.013	1.365	.227
	Within Groups	872.789	592	1.474		
	Total	884.864	598			
Role of heads of departments	Between Groups	22.704	6	3.784	4.519	.000
	Within Groups	453.016	541	.837		
	Total	475.720	547			
Value of heads of departments	Between Groups	29.639	6	4.940	4.799	.000
	Within Groups	556.853	541	1.029		
	Total	586.492	547			

Table 4.47 shows that the main difference between subjects with regard to the role of head teachers is between English teachers (M=2.86) and Social Studies teachers (M=3.39). This indicates that English teachers see head teachers as less involved in their evaluation than Social Studies teachers do. English teachers (M=2.62) rate the value of the head teachers' role lower than do Social Studies teachers (M=3.37), Arabic teachers (M=3.17), and Islamic Studies teachers (M=3.09).

The greatest mean difference between subjects with regard to the role of heads of departments is between Computer Science teachers (M=2.87), who see heads of departments as less involved than Social Studies teachers (M=3.94), Arabic teachers (M=3.62), and Islamic Studies teachers (M=3.59).

The greatest mean difference between subjects with regard to rating the value of heads of departments is found between Computer Science teachers (M=2.88), who rate their value as less than do Arabic teachers (M=3.69), Islamic teachers (M=3.70), and Social Studies teachers (M=3.92).

Table 4.47: Tukey's HSD post hoc test for the involvement of evaluators, subjects

Involvement of evaluators	Grouping variable	N	Mean	Mean Difference					
				English	Islamic	Science	Math	Social	Computer
Role of head teachers	Arabic	132	3.21	.352	.023	.001	.240	.175	.142
	English	93	2.86		.375	.353	.112	.528*	.209
	Islamic	111	3.24			.022	.263	.152	.166
	Science	80	3.22				.241	.174	.143
	Math	98	2.97					.415	.097
	Social	59	3.39						.318
	Computer	26	3.07						
Value of head teachers	Arabic	132	3.17	.552*	.089	.170	.315	.193	.025
	English	93	2.62		.462*	.381	.236	.745*	.526
	Islamic	111	3.09			.081	.226	.282	.063
	Science	80	3.00				.144	.364	.145
	Math	98	2.86					.508	.289
	Social	59	3.37						.219
	Computer	26	3.15						
Role of heads of departments	Arabic	127	3.62	.150	.032	.196	.225	.310	.754*
	English	87	3.47		.118	.045	.074	.461	.603
	Islamic	101	3.59			.164	.192	.342	.722*
	Science	70	3.43				.028	.506*	.558
	Math	89	3.40					.535*	.529
	Social	50	3.94						1.06*
	Computer	24	2.87						
Value of heads of departments	Arabic	127	3.69	.310	.019	.385	.192	.236	.801*
	English	87	3.37		.330	.074	.118	.547*	.490
	Islamic	101	3.70			.404	.211	.217	.820*
	Science	70	3.30				.193	.621	.415
	Math	89	3.49					.428	.609
	Social	50	3.92						1.03*
	Computer	24	2.88						

4.3.4 The extent to which the current system supports teachers

4.3.4.1 Gender

Table 4.48 shows that there is no statistically significant difference between female and male teachers regarding their perception of how well the system supports them in their performance development, where the p-value for each scale is p >.05. However, there is a statistical difference between gender with regard to the extent to which the system supports the awarding of promotions and rewards. The p-value is .0001 and reflects that female teachers (M=3.50) believe that the current system supports them in being granted promotions and rewards more than male teachers (M=3.09).

Table 4.48: The independent samples t-test of gender for the system supports teachers.

The system supports	Grouping variable	N	Mean	Std. Deviation	Std. Error Mean	p-value
Development of performance	Male	210	2.94	1.12	.077	.442
	Female	389	2.87	1.20	.059	
Promotions and rewards	Male	210	3.09	1.16	.080	.000
	Female	389	3.50	1.06	.053	

4.3.4.2 Educational districts

The results of the ANOVA test in Table 4.49 show that there are no significant differences between educational districts with regard to the extent to which the system supports both development of performance, and the awarding of promotions and rewards scales. Here, the p-values for all scales are greater than .05.

Table 4.49: ANOVA results of educational districts for the system supports teachers.

The system supports	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Development of performance	Between Groups	.516	2	.258	.191	.826
	Within Groups	803.703	596	1.348		
	Total	804.219	598			
Promotions and rewards	Between Groups	5.752	2	2.876	2.324	.099
	Within Groups	737.636	596	1.238		
	Total	743.387	598			

4.3.4.3 Years of teaching experience

Table 4.50 shows that statistically significant differences between teachers' perceptions of the extent to which the system supports them in developing their performance according to levels of experience; the p-value for scale is p=.0001. By contrast, no significant differences are found in perceptions of the extent to which the system supports the

awarding of promotions and rewards between the groups' experience levels, where the p-value is p >.05.

Table 4.50: ANOVA results of experience groups for the system supports teachers.

The system supports	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Development of performance	Between Groups	34.718	2	17.359	13.445	.000
	Within Groups	769.501	596	1.291		
	Total	804.219	598			
Promotions and rewards	Between Groups	4.017	2	2.008	1.619	.199
	Within Groups	739.370	596	1.241		
	Total	743.387	598			

Table 4.51 shows that teachers with more experience believe that the current system supports them in developing their performance when compared to teachers with less experience. The greatest mean differences are between teachers with less than 10 years' experience (M=2.67) and teachers with more than 20 years of experience (M=3.44).

Table 4.51: Tukey's HSD post hoc test for the system supports teachers, experience groups

The system supports	Grouping variable	N	Mean	Mean Difference	
				Between 10 and 20	More than 20 years
Development of performance	Less than 10 years	285	2.67	.355*	.767*
	Between 10 and 20	260	3.03		.411*
	More than 20 years	54	3.44		

4.3.4.4 Subjects

The results of the ANOVA test in Table 4.52 show no significant difference between subjects with regard to the extent to which the system supports both development in performance, and the awarding of promotions and rewards. Here, each scale is p >0.05.

Table 4.52: ANOVA results of subjects for the system supports teachers.

The system supports	Grouping variable	Sum of Squares	df	Mean Square	F	p-value
Development of performance	Between Groups	12.356	6	2.059	1.540	.163
	Within Groups	791.863	592	1.338		
	Total	804.219	598			
Promotions and rewards	Between Groups	7.017	6	1.169	.940	.465
	Within Groups	736.371	592	1.244		
	Total	743.387	598			

4.3.5 Summary of inferential statistics

Table 4.53 organises the differences in terms of each item and scale, with regard to four variables: gender, educational districts, experience, and subjects. It can be seen that the most statistically significant differences with regard to the purpose, either actual or desired, are found between educational districts and groups of teachers with different amounts of teaching experience.

With regard to the statistical differences in the tools that are used, these are found between all variables. As for the tools that should be used, the statistical differences are found between educational districts, experience groups and subjects taught. With regard to the statistical differences regarding the involvement of the evaluators, statistical differences are found between educational districts, experience groups and subjects.

When it comes to the extent to which the system supports teacher in developing their performance, the differences occur primarily according to the number of years of teaching experience. As for the extent to which the system supports the awarding of promotions and rewards, the main difference can be found according to gender.

The implications of the differences found in teachers' responses through the analysis of the quantitative data will be discussed in chapter 8.

Table 4.53: Summary of inferential statistics

		Gender	Educational districts	Experience	Subjects
The actual purposes	Professional development		*	**	
	Determining performance				
	Sanction and rewards		*		
The desired purposes	Professional development		*		
	Determining performance			*	
	Sanction and rewards				
Tools: are used	Observation	**			
	Students' achievements		**	**	
	Self-evaluation		*	**	**
	Peer evaluation 'formative'			**	
	Students' evaluation		**	*	
	Teachers' portfolios				*
Tools: should be used	Observation		**		*
	Students' achievements		**		**
	Self-evaluation		**		*
	Peer-evaluation		**	*	*
	Students' evaluation		**		**
	Teachers' portfolios		**	*	**
The role of evaluators	Head teachers		**		*
	Inspectors			**	
	Heads of departments				**
Rating the value of the evaluators' role	Head teachers		*		**
	Inspectors			**	
	Heads of departments				**
The system supports	Development of performance			**	
	Promotions and rewards	**			

*p < 0.05 // **p < 0.01

Chapter Five: The Perspectives of Head Teachers and Inspectors on the Current Teacher Evaluation System

5.1 Introduction

As stated in *Chapter Three*, inspectors' and head teachers' perspectives on the current teacher evaluation system were collected via semi-structured interviews.

In the analysis, data collected from the interviews were divided into three main themes, which were further divided into sub-themes. The main themes are: the purposes of teacher evaluation (the actual and desired purposes); the tools of teacher evaluation (tools that are used and that participants think should be used); and the evaluator's role in teacher evaluation. Each of these themes will be discussed below. Outcomes of the interviews with head teachers will be considered first, followed by the outcomes of the interviews with inspectors. Table (5.1) lists the participants' (fictive) names, gender and years of teaching experience.

Table 5.1: Name of interviewees and their teaching experience

Head Teacher	Gender	Experience
Awatf	F	29 years
Ghadeer	F	29 years
Hasah	F	23 years
Loui	M	29 years
Maharb	M	35 years
Mariam	F	34 years
Noriah	F	27 years
Shafah	F	33 years
Waleed	M	22 years
Inspector	**Gender**	**Experience**
Abdualkreem	M	16 years
Ali	M	19 years
Alia	F	17 years
Fahad	M	25 years
Hada	F	21 years
Hadel	F	17 years
Mohammed	M	21 years
Mubarak	M	28 years
Nawaf	M	20 years
Noor	F	26 years
Salwa	F	30 years
Wafa	F	25 years

5.2 Head teachers

This section presents an analysis of the data collected from interviews with nine head teachers in primary schools. The analysis is based on an examination of the answers that they held in common and those that differed.

5.2.1 The purposes of teacher evaluation

5.2.1.1 The actual purposes of teacher evaluation

Head teachers were asked about the purposes served by the current system. The common answers given by head teachers is that the current teacher evaluation system is concentrated on summative purposes, namely, all head teachers reported that they and other evaluators evaluate teachers in order to annually determine the individual teacher's performance as weak, good, very good, or outstanding. Head teachers submit final reports about the individual teacher's performance to the MOE *on behalf of* CSC at the end of the school year, based upon which decisions are made regarding a teacher's promotion, salary scale, and annual bonus to be given as a reward, if appropriate, or sanctions to be made. In extremis, this might mean transferring the teacher to a non-teaching role or lead to their dismissal. For example, Waleed explained:

> *I work with inspectors and heads of departments to identify individual teacher performance each year and give a rating. Then, based on our decision the MOE makes decisions and thus teachers will be awarded increases in the salary scale, promotions, annual bonus, or either dismissal or referral to a non-teaching job*.

Heads teacher were asked how the MOE uses the teacher evaluation reports in imposing sanctions and giving rewards. All nine head teachers reported that a teacher has to achieve an outstanding grade to gain a promotion or a reward. For example, Maharb stated "*the MOE started applying new regulations to the Teacher Salary Scale no. 28/2011. Teachers have to achieve an outstanding performance in order to obtain salary increases*". Mariam also explained that:

> *Based on the final teacher evaluation report that is outstanding, I and other heads of departments or inspectors have awarded promotions (to become head of department then inspector or head teacher). Furthermore, there is an annual*

bonus given to all teachers who achieved an outstanding performance based on their teacher evaluation.

If a teacher's performance is evaluated as weak, on the other hand, sanction will be taken against that teacher. For example, Ghadeer stated "*if a teacher had been evaluated as weak during their evaluation - three consecutive years for Kuwaiti teachers or one year for foreign teachers - that teacher will be dismissed or forced to leave teaching for other administrative work*".

Other factors emerged as important sub-themes that influence the use of current teacher evaluation system for professional development: lack of openness; lack of motivation with regards professional development; focus on non-teaching duties in the evaluation; lack of collaboration between head teacher and inspector

First, the regulations regarding teacher evaluation state that teachers are not informed about their evaluation as the final report must be kept confidential. Eight head teachers mentioned that keeping the reports confidential hinders the use of the current system of teacher evaluation in improving and developing teachers. For example, Waleed asked, *"How can I support teachers to develop or improve their performance with regulations that state that teacher evaluation must be kept confidential from the teachers themselves?*" In order to circumvent this rule, some head teachers provide specific feedback after observation that includes advice and suggestions to teachers about their performance throughout the school year. Shafah, for example, stated:

I provide advice, suggestions, and highlight the strengths and weaknesses of the teacher's performance immediately after classroom observation, but it is still a personal initiative; but in the end, I cannot do that for teachers, as their evaluation details must be kept confidential.

Ghadeer felt head teachers should provide feedback to teachers after observation; however, this simply consists of a standardised observation checklist of what head teachers have observed inside the classroom. So she also helps teachers by providing advice, recommendations, and information that is not included in the official checklist. Meanwhile, Hasah believes that providing feedback to teachers cannot replace the

importance of the final reports about their performance: "*feedback is an attempt, but to be honest, does not equal our ambition or is the best way of improving teaching as it would be if teachers could receive the final report of their performance*". Awatf also pointed out that feedback tends to focus on what is observed inside the classroom, and thus makes little improvement regarding pedagogy, classroom management or questioning techniques.

The second sub-theme to emerge is the lack of motivation to attend further training and workshops despite inspection departments providing training courses and workshops for teachers on behalf of the educational districts. All head teachers stated that teachers are not obliged to attend these courses. However, weak teachers (teachers whose performance has been evaluated as weak) are obliged to attend training courses designed by the inspection departments or other departments in educational districts, or other schools, to improve their performance. Inspectors are expected to create a training plan and enhance supervision for these teachers before making the decision to either dismiss them or refer them if they do not improve. Mariam stated:

> *We cannot force outstanding or good teachers to improve or develop their performance and attend training courses, only weak teachers are obliged to improve their performance and attend either courses or training to avoid dismissal or transfer to a non-teaching job.*

Shafah also confirmed "*... with good and outstanding teachers, evaluators choose which of them attend courses, but neither schools nor inspectors can oblige those teachers to attend because course attendance is not accounted for in their evaluation*". Loui pointed out another reason for teachers not attending these courses: most courses are conducted during school hours and therefore few teachers are able to fit attendance of these courses in with their daily responsibilities and many teaching sessions per day.

With regards to non-teaching duties, three head teachers shared the common view that some head teachers or inspectors concentrate more on non-teaching duties than on teaching or other parts of teacher performance in the evaluation. Loui, Waleed and Mariam all stated that some evaluators only focus on the degree to which a teacher has cooperated with school management and their involvement in organising school activities when carrying out a teacher evaluation. This leads some teachers to focus on organising school activities in

order to obtain an outstanding performance grade, more than seeking to improve or develop their teaching performance. For example, Mariam stated:

> *Some head teachers focus on what activities teachers have organised, to show up their own schools to other school and district. This leads teachers to not care about learning or teaching if they do some activities for school, in order to be rated as outstanding.*

The final sub-theme to emerge was the lack of collaboration between head teacher and inspector. According to Loui, there are many problems associated with inspectors visiting two to three times a year and trying to impose their views on teacher performance rather than collaborating with head teachers and thus the feedback provided can often be inconsistent with regards to what exactly needs to be improved or developed in order to get an outstanding grade.

5.2.1.2 Head teachers' desired purposes for teacher evaluation

Head teachers were asked about what purposes should *ideally* guide teacher evaluation. On this question, all head teachers concurred that teacher evaluation should comprise both summative purposes and professional development. The shared view was that teacher evaluation should be used to make decisions about underperforming teachers, and give promotions to teachers as recognition. Teacher evaluation should develop performance, and thus improve students' learning. For example, Awatf explained, "*Professional development is needed to draw the map for teachers to improve and develop their practice*". Ghadeer also stated:

> *I would use teacher evaluation to support teachers with skills to teach, encourage teachers to develop. This will help and lead to high-quality teaching, and thus improve students' learning, since teachers are one of the most important principles regarding the quality of education*.

Furthermore, as Loui explained, "*summative evaluation is needed in teacher evaluation to protect the rights of teachers in awarding rewards and promotions*". Maharb added "*Sanctions and rewards could encourage teachers to do a great job to obtain promotions*

and avoid any sanction. With some teachers, if there is no sanction they may not be hard-working".

Head teachers were next asked what was needed to achieve these purposes. The following points were raised.

First, seven head teachers returned to the issue of the final evaluation and thought the final detailed report should be given to the individual teacher. A common argument was that teachers would become more aware of what they have achieved and what they need to improve upon, while encouraging outstanding teachers to maintain that level. For example, Noriah argued that "*if teachers know what they have achieved and need to improve on, they will think and start to self-reflect on how to become outstanding teachers"*. Shafah also stated that:

> *The final detailed evaluation report will be a mirror for teachers: teachers will know what they've achieved; what they need to improve or develop, etc. For outstanding teachers, they will be happy after they see their achievement at the end, then it can motivate them to continue to do such great work in the next year.*

On the other hand, Loui considered a different approach to providing the final report:

> *I prefer only to provide the criteria of teacher performance with comments - no score ratings. 'Scores must be kept confidential'. I think providing the report in detail will cause problems for head teachers and their teachers. For example, a teacher will ask why 70 not 90, or some teachers will not collaborate with school management in the next year as they were given a low score*.

By contrast, Maharb prefers providing only a mid-year report that includes details (with a score) about a teacher's performance after the first semester, and believes that the reports at the end of the school year including further details and a score, should be kept confidential. Asked if he prefers this way because of potential problems that he would face by providing the final report about a teacher's performance, he responded:

This is my decision about the teacher, and I prefer to keep that confidential. It is not about making problems for me with teachers. I would not like teachers to know what I am saying about their performance. Teachers will know if they did not achieve an outstanding performance by their opportunities to be promoted, their salary scale and their annual bonus.

Other head teachers were also asked whether reforming the system would cause problems, such as less collaboration with school management or personal problems when teacher do not get the score they are expecting. They pointed out that the head teachers would not have problems providing a final report to teachers since everything would be made clearer by showing them their strengths and weaknesses. Hasah thought that detailed teacher reports might lead to initial problems with teachers who do not accept criticism, but that in time, teachers would accept the process and that when the report shows teachers their strengths and weaknesses, the purpose of the exercise would become evident to them.

With regards to training courses and professional development activities, Shafah suggested that these should be organised by inspection departments in the educational districts based on the final reports of the teacher evaluation, so as to meet the teachers' needs, "*leading to an improvement in terms of weaknesses, building on their strengths, and keeping teachers up-to-date*". She added "*Sometimes our subject departments in school receive invitations from the inspection departments for courses on how to use a computer, yet many teachers in our school have (International Computer Driving Licence) and use smart devices (iPads) in their classes*". In other words, courses are often not appropriate for many teachers or do not meet the needs of many teachers.

In Waleed's view, there should be financial support for schools to organise such workshops and training for the whole school and these could be designed by heads of departments or head teachers, based on the teachers' needs:

With financial support, school staff will be able to keep teachers up-to-date via some courses in school that are considered suitable for the teachers and teachers will be able to attend as it is in our school, and thus our own courses may more easily meet our teachers' needs.

Noriah agreed "*Schools need financial support to organise training courses and schools should also have the freedom to collaborate with some private agencies. These are able to assist us, as school staff, to organise different types of courses*". She explained that as they work in public schools, head teachers are not able to collaborate with private agencies or accept donations from agencies without permission from the MOE, which can take a long time to obtain. Therefore, school staff "*depend on support from academic staff at Kuwait University as this is a government university, and schools just need permission from the educational activities director in the district*".

There was also a call for more explicit and strict sanctions for teachers. Waleed thought these should potentially lead to dismissal when a teacher does not perform well, and that other rewards should be available in addition to money or promotions which lead indirectly to professional development, such as scholarships and travel bursaries to attend conferences: "*those are needed to motivate teachers to perform well and increase the commitment to work*". When asked about the meaning of explicit sanctions, he explained "*Sanctions must be applied after teachers have been given one or two more years to improve. The MOE should not postpone a decision because of lack of teachers in particular subjects such as Mathematics or Arabic*".

5.2.2 The tools of teacher evaluation

5.2.2.1 The tools of teacher evaluation that are currently being used

Head teachers were asked about the tools currently being used to evaluate teachers' performance.

Observations and teacher portfolios are used, with all nine head teachers in this study stating that observation is always used in teacher evaluation. For example, Hasah stated " *I observe teachers' practice inside the classroom three times in one semester*". They also commonly take into account teacher's portfolios to get information about a teacher's activities and achievements during the school year. For example, Maharb stated "*I use portfolios at the end of the first semester and at the end of school year".*

Shafah, Ghadeer, Awatf, Hasah, and Noriah indicated that they have occasionally used student achievement and self-evaluation as a personal initiative in their teacher evaluation. Shafah stated "*I ask teacher to give comments about themselves in my evaluation files, this*

is my decision. Teachers' opinions about their own performance can show me some aspects I did not cover, which can help to make the final judgment". Awatf also explained:

> *Student achievement is useful to reflect teacher performance. During the school year, I follow up with student achievement and I focus on teachers with low student achievement to see if the problem is with the teacher or students. Then I try to help or make my judgment. This approach is not applied in many schools. I apply it as a personal initiative.*

With regards to peer evaluation, head teachers reflected that they do not mention peer evaluation for formative or summative evaluation. Shafah were asked about peer evaluation and she stated "*I do not ask head of department or teachers whether or not teachers evaluate each other, to write that in their final report as a collaboration effort*". Noriah also stated "*this is the role of departments heads who can look at peer evaluation. They are responsible for seeing collaboration among teachers in their departments whereas I look at collaboration in the school"*

Asked why they used observation and portfolios more than other tools, the reasons head teachers' gave can be divided into two sub-themes:

Firstly, the policy on teacher evaluation is a main influence. All nine head teachers pointed out that the policy states that individual teachers have to be evaluated inside the classroom. All evaluators have to complete a standardised observation checklist several times throughout the first and second semesters. These checklists help them to write the final report about the teacher's performance throughout the school year. They also explained that based on certain criteria, they look at a teacher's portfolio to help them to evaluate a teacher's activities and achievements as required from evaluators (See the criteria of teacher evaluation in appendix 1). As Hasah stated, "*observation is a legal tool to see teacher inside the classroom, but the teacher's portfolio is a tool that helps us to measure a teacher's participation in school activities during the school year*". Maharb also stated "*Portfolio is used to reflect what teachers have done in terms of activities inside the classroom or in school*". Ghadeer and Shafah commented that it is difficult to remember what teachers have done during the school year with regard to their activities or achievements, so a teacher's portfolio helps them to fill out the teacher's evaluation report.

Secondly, teachers' collaboration with evaluators is an important factor: Loui and Waleed believe there is a lack of collaboration from teachers when certain tools are used to evaluate them, particularly tools that are not explicitly mentioned in the teacher evaluation policy. Loui stated "*Teachers argue with me; for example, when I look at students' achievements or conduct interviewing students and link that to their final performance report*". While Waleed found he could not

> *support teachers to do, e.g. peer evaluation, either formative or summative, or students' evaluation, when teachers think this is an illegal tool to use in their evaluation. Especially since many teachers believe that teacher evaluation is used to control and judge them in order to make decision, not to help them*.

5.2.2.2 The tools of teacher evaluation that ideally should be used

When asked about the tools of teacher evaluation that they ideally would like to use, all head teachers, it seems, agreed that observations and portfolios are necessary tools to use, while eight thought that evaluation should be based on using a wider range of tools, i.e. student achievement, peer and self-evaluation. In this regard, Waleed argued that such tools should be included in the policy of teacher evaluation and used by all evaluators: "*do not leave the use of multiple tools to personal initiative. It should be included in the policy of teacher evaluation*".

However, one head teacher, Maharb, disagreed with the majority, stating that "*observation and the portfolio are more than enough to distinguish between outstanding teachers and others, in order to a make judgment about a teacher's performance*".

A range of arguments was made among those who supported using different tools in teacher evaluation. Shafah stated "*using different tools of teacher evaluation leads to a prevention of emotion when it comes to judging a teacher's performance. Furthermore, different tools are better than only one person visiting her inside the classroom and determining performance".* Along similar lines, Waleed stated, "*using several tools in individual teacher evaluation will reduce the subjectivity when evaluating teachers and make fairer judgments than one tool*". Loui pointed out, *"using several tools will assist us*

to collect reliable data about teachers' performance which then can help us to make valid judgments" and later added "*Using different tools when evaluating teachers would be helpful for teachers. They will agree with their final reports due to the inclusion of different tools in the evaluation, not simply observation and looking at their portfolio as happens now".*

There was general agreement that peer evaluation as a tool should be used for formative purposes, in order to facilitate the exchange of experiences between teachers. Waleed stated "*our school has teachers with a lot of experience and other outstanding teachers who have gained prizes in teaching in Kuwait and Arabian Gulf Countries. They can facilitate development of other teachers through using peer evaluation for formative purpose*". The reason given for why peer evaluation should be formative and not summative is that by using it for formative purposes, invalid judgements can be avoided and teachers are encouraged to help rather than judge one another. For example Awatf stated, "*By using it formatively, the teacher does not have to fear peer evaluation, thus they will be encouraged to evaluate each other*". Hasah thinks that "*some teachers have personal problems with each other, and with summative evaluation they may try to judge each other unfairly*".

With regard to student evaluations, the head teachers have contrasting views. Awatf and Hasah agree that students are *too emotional* and cannot adequately determine a teacher's performance. Hasah, for example, stated:

> *Many students hate teacher who gives them a lot of homework, or strict teachers. When I ask them about teachers, they will give answers based on their hate or love. I mean students will say something good when they love the teachers, if they do not like the teachers they will say something untrue*.

However, six of nine head teachers agreed that using student evaluations can provide valuable information about teachers regarding teaching and activities inside the classroom. For example, Waleed stated "*the evaluator can ask students questions such as do you enjoy a particular teacher's session and why. Students will talk and I will find out some interesting information*". On the other hand, Noriah argued that "*I do not need students determining teaching. There are many things included in a teacher evaluation, such as*

giving respect to students, and dealing with them, and that can be seen without the students' knowledge". Shafah's position was that:

> *Nothing prevents us from including students' evaluations in our teacher evaluation. We can do this either by a simple questionnaire that includes a happy face to answer indirect questions or by interviewing students for a few minutes with indirect question. Then an evaluator will analyse the students' responses. The evaluators can recognise if students just like the teacher or if the teacher is doing well*.

5.2.3 The involvement of head teachers as evaluators in teacher evaluation

Asked about the role of evaluators, all nine head teachers concurred that they provide written feedback to teachers after each observation so the teachers know what head teachers have discovered based on the observation. Eight head teachers also stated that they always spoke with individual teachers following their observations as holding a discussion is an opportunity, in their view, to explain their comments to the teacher. For example, Loui stated "*I discuss with teachers after observation to explain to the teacher my feedback and what I observed".* Maharb, on the other hand, stated that he did not speak with all individual teachers after their observations because they (he and the teachers) had too many significant responsibilities in a school day.

In terms of their role and the value of their role as evaluators, based on the data from the head teachers, there are four factors that may come into play:

The first is a lack of subject knowledge. Waleed, Shafah, Noriah and Awatf indicated that they found their lack of subject knowledge to be a problem when they evaluated teachers from different subjects. For example, some head teachers were Arabic teachers and were evaluating maths or English teachers. However, they were able to evaluate the teachers in terms of transferrable knowledge in pedagogies and classroom management. They also asked the heads of departments to attend in order to make up for their lack of subject knowledge and ask them about any issues that they perceived. Waleed explained that:

Since the head of department is more specialised than me, there is no problem in using their specialist skill to help me inside the classroom to evaluate teachers in terms of the content and some part of the teaching in order to make a fair and accurate judgment.

Second, all of the head teachers indicated that they have significant responsibilities within their schools and in managing them. They have to observe individual teachers each semester and provide them with feedback, each teacher being observed on average between 2 and 4 times each school year. This is multiplied by between 70 and 100 teachers in every school. They do not have as much time as the heads of departments and they believe this may affect the value of their role compared to the heads of departments who only have a few teachers to evaluate.

Third, an absent teacher may also affect the head teacher's role and its value, as Hasah explained:

I organise a timetable for every teacher, but I encounter difficulties when a teacher is absent on her visit day, since I cannot slot her in on another day since my schedule is already full. Therefore, this problem affects the evaluation, as though I will try to conduct another visit, sometimes I cannot conduct a complete observation session, or even discuss the observation with her afterwards.

The fourth issue is the lack of training. Shafah, Awatf, and Mariam share the view that lack of training for head teachers affects the value of their role. Head teachers should have opportunities to develop their knowledge in the field of education and evaluation and stay on top of other innovations in education as well.

5.3 Inspectors

This section presents the themes identified in interview data with twelve inspectors in three educational districts with reference to similarities and differences found between their views and those of the head teachers.

5.3.1 The purposes of teacher evaluation

5.3.1.1 The actual purposes of teacher evaluation

As with the head teachers, inspectors were asked about the purposes of the current evaluation system. All except one inspector defined its purpose as being to annually determine individual teacher performance, and helping the MOE to make decisions regarding promotions, salary increases or sanctions. For example, Mubarak explained that:

> *Teacher evaluation is used for determining a teacher's performance by reflecting a teacher's knowledge, subject matter, the extent to which they follow the national curriculum, all functions as a teacher and giving scores that reflect individual performance. Furthermore, the outcomes of teacher evaluation are also used by the MOE for promotions, salary increases or dismissal or transferring individuals to non-teaching jobs.*

However, Hadel pointed out that while teacher evaluation is used for sanctions for all teachers, it is only used for rewards with Kuwaiti teachers:

> *Teacher evaluation is used to award rewards for Kuwaiti teachers, as foreign teachers do not get the same chance for promotions. They only get a promotion (to be a head of department) if there is no Kuwaiti in their subject department*.

The sub-themes to emerge from the inspectors' perspectives revolve around explanations for why the purposes of the current teacher evaluation concentrate on summative. The sub-themes are as follows:

First, eleven inspectors concurred with the teachers regarding the current system's failure to promote professional development due to the final report not being formally shared with the teachers. As Mubarak, Salwa and Wafa pointed out, by keeping teacher evaluation reports confidential, the teacher cannot know their progress or weaknesses at the end of

school year and therefore lack the information that would allow them to judge their own efforts to improve and develop. Moreover, as Ali and Fahad explained, because the details of the final reports are unknown, teachers lack awareness of the criteria that determine whether or not their teaching is adequate and what they must do (or avoid) in order to become better teachers. For example, Ali stated:

> *Teachers will only learn from feedback that shows his/her weaknesses or errors. Teachers cannot compare his/her performance overall to what effective teaching is based on the criteria of the teacher evaluation in order to avoid some things or improve on others*.

As Nawaf pointed out, teachers may well work to improve and develop their own skills initially but "*They work in the first, second, third…. 10 years, but then they will feel bored when they no longer know exactly what they have achieved. Then maybe they will stop trying to improve and develop, and no longer care*".

Furthermore, Abdualkreem indicated that not knowing the contents of the final report affects inspectors too. A new inspector will have no idea about the history of a teacher and she/he has to spend time identifying the teacher's strengths and weaknesses rather than being able to use the previous inspector's work to help teacher. Mubarak, a head inspector, was asked whether he could provide the previous teachers' performance reports to new inspectors so they could learn from them, to which he responded: **"***I cannot provide my colleagues with the reports of teachers for their teacher in new schools as I have to keep these confidential with the head teachers and with the inspector who evaluated teachers in a particular school*".

Second, inspectors share the head teachers' opinion regarding the obligation for teachers to attend training courses and workshops. Inspectors were asked about teachers' attendance at departmental courses or workshops, which are held at schools and designed by the school in collaboration with the inspection departments. All inspectors explained that when a teacher is evaluated as being weak, they receive intensive evaluation throughout the following year. As part of this, that teacher must attend training courses and workshops and their attendance is recorded in their report as evidence that they have made an effort to

improve. If they do not improve by the end of the following year, they may be dismissed and forced to leave teaching to pursue other administrative work within the MOE.

On the other hand, inspectors stated that there is no obligation for other teachers to listen to what an inspector suggests they should do in order to develop. For example, Fahad states, "*teachers sometimes attend as a kind of collaboration with the inspection department*". Wafa believes one obstacle to attending such courses is the teachers' fear of how others will perceive them:

> *Very good or outstanding teachers do not attend our courses because they do not want others to think that they are 'weak' teachers, since the weak teachers are sort of forced to attend. They do not want to be lumped in with the 'wrong' crowd.*

Third, Noor pointed out another obstacle to professional development within the current system. She stated "*some evaluators make teachers feel teacher evaluation is system to control them and given sanction or promotion*". She went on to say that some inspectors, when visiting teachers inside the classroom, only want to see mistakes and use teacher evaluation as a way to control teacher "*you have to do this…, to be head of department…*".

5.3.1.2 The desired purposes of teacher evaluation

The data from the interviews show that, like the head teachers, all inspectors believe teacher evaluation should be used for summative purposes and professional development. With regards to the importance of the summative purposes, Alia explained that "*... is needed to provide information about teacher performance that can be used by the MOE to reflect the quality of teaching in public schools or by the CSC when publishing reports about employee performance in the public sector".* All inspectors agreed that teacher evaluation is necessary as it acts as a means of recognising their hard work. For example, Abdualkreem stated "… *not all teachers deserve to be teachers. Through teacher evaluation, the MOE can distinguish between teachers who are hard-working and other teachers*". Fahad stated, "*the MOE can organise promotions, and a salary scale for teachers based on their performance*". Hada thinks that "*by sanction and reward, teachers may be motivated to work hard and improve their performance to be outstanding teachers and gain promotions and salary increases*".

With regards to the importance of professional development in teacher evaluation, inspectors believe this is necessary for teacher themselves and students' learning. For example, Ali stated "*we (Inspectors, teachers, head teachers, all staff in the MOE) work for students and their learning. Promoting professional development is needed in order to provide high quality learning for students*". Job satisfaction was also mentioned in relation to professional development. Mohammed stated:

> *Professional development is important for job satisfaction. Teachers do not need only a good salary to feel contentment about their work. They need to improve and develop to meet high quality of teaching to feel high level of contentment about their performance then their work as teacher.*

Asked what was needed to achieve these purposes, the following points were raised:

First, as the head teachers, all the inspectors thought the final evaluation report and all its details should be made available to teachers, as this would enable teachers to work on their weakness and strengths, understand what constitutes effective teaching, and thus know how to deliver the best performance that they can. For example, Fahad stated "*Teachers will know everything about their performance from the final reports of individual evaluation, assisting them to improve themselves and to renew their personal and professional growth"*. Ali stated "*Teachers will know what's expected from them, and what effective teaching is… to do their best*". Noor added that in fact teachers may know about their report informally but it is important for "*teachers to know of their progress in a formal way*". Nawaf argued "*there is no reason to keep the final reports confidential. If the final reports may make a problem for evaluators, I do not care about some problems with teachers if I state the truth"*. Furthermore, Abdualkreem suggested that the CSC *"should look at teachers as teachers, not as employees in other ministries and change Regulation No. 36/2006 so that individual teachers can know their evaluation reports*".

Second, by attending courses, inspectors believe that teachers will be kept up-to-date and will have a better understanding about how to improve or develop their performance. As Salwa stated, "*Teachers will have the keys to develop their teaching"*. However, the problem is the low attendance of teachers on these courses. Hence, Mubarak suggested that

a "*record of training courses in teacher evaluation will encourage teachers to attend our courses, and a record of training will help evaluators to appreciate teachers' efforts who attend our courses*". Alia also suggested keeping a record of attendees:

> *We (inspectors) invite some teachers to our training courses but they do not usually attend. If a record of training courses is part of the teacher evaluation, me and the other evaluators can judge teachers who do not have an acceptable reason for non- attendance.*

Third, Noor suggested that teachers should be know in advance when an evaluator plans to come for formative evaluation, and that teacher evaluation should not be threatening. For example, when an evaluator observes a teacher for formative purposes, the evaluator should tell the teacher " I am here today just to see how good you are, and give you some advice"

The fourth issue raised was that of rewarding foreign teachers and not just Kuwaiti teachers. One inspector, Hadel, suggested greater rewards for all teachers *including* foreign teachers "*I am not talking about salaries, as there are a lot of regulations within CSC on behalf of the Kuwaiti Government, but I am talking about the opportunities for promotions*". She believes that:

> *If they are doing the same thing as Kuwaiti teachers, then they should have access to the same opportunities to become department heads, inspectors, or deputy head teachers then head teachers, instead of preferring Kuwaiti teachers for these positions and giving priority to Kuwaitis.*

5.3.2 The tools of teacher evaluation

5.3.2.1 The tools of teacher evaluation that are currently being used

In terms of the tools inspectors used to evaluate a teacher's performance, interview data shows that observation is the most commonly used. Only two inspectors also take into account a teacher's portfolio and student achievement as well as observation before drawing a final picture or making a final judgment about a teacher. During their visits to observe teachers, Mohammed and Mubarak look at the teacher's portfolio and the student

achievement to complete the impression that they have about the teacher's performance. For example, Mohammed stated

> *I have 200 teachers to evaluate in different levels of schools every school year. So, I cannot visit all of them in 45 minutes* [length of session]. *I visit them for between 15 and 30 minutes, three times a year. Therefore, I cannot make an accurate judgment by observation alone, so I look at the portfolio and students' achievements just to confirm my opinion about the teacher.*

Different reasons were given by those inspectors who only used observation to evaluate teachers. Firstly there is a fear of taking personal initiative, based on a reluctance to do anything that is not explicitly mentioned in the teacher evaluation policy. For example, Ali stated: "*I respect the regulations of the MOE and I am committed to the policy of teacher evaluation. Thereby, I am afraid to use anything that comes under my personal initiative that might be used against me*". Hadel shared this view, explaining that as an inspector she does not use a range of tools because the head inspector will reprimand her if she does not follow the policy of teacher evaluation.

For Abdualkreem, observation gives inspectors enough of a picture about a teacher's performance to be able to make a judgment. He stated: "*I can look at different aspects of performance such as teachers' knowledge of content, pedagogies, behaviours, students' engagement, and classroom management*". Noor also stated "*observation is more than enough to make a decision about a teacher"*. She added that the inspector is not in school to witness the events recorded in the teacher's portfolio, for example and therefore:

> *As an external evaluator, I do not trust the teachers' portfolio. A teacher's portfolio is only paperwork (it is like ink on paper) if I am not with the teachers step by step. This applies to other tools such as self-evaluation*.

5.3.2.2 The tools of teacher evaluation that ideally should be used

As with the head teachers, ten of the inspectors support using a number of different tools to evaluate teachers in order to make more accurate judgments about their performance, i.e. teacher portfolios, student achievement, student evaluation and self-evaluation. According

to Hadel "*various tools could look at teacher performance in different ways. These assist us in making judgments about teacher performance during the school year inside and outside the classroom, so why not use them?*" Ali concurred "*By using these tools, teacher evaluation will be more accurate in reflecting teacher performance inside and outside the classroom*". Hada pointed out that using different tools leads inspectors to make more accurate and fairer judgments on individual teacher performances than by simply depending on one tool, such as observation. Furthermore, Mohammed believes these tools not only help evaluators but teachers too: "*a teacher will have better information about their performance than when the only tool is observation, and thus better information will facilitate development*"

While observations help inspector to determine teacher's practice inside the classroom, student achievement help inspectors to see what the teacher has done with students' learning. For Wafa, "*students' achievement is the outcome of the teacher's practice, and the evaluators see it as a reflection of teacher performance*". Student evaluation gives inspectors information about students experience with their teachers. For example, Nawaf commented that "*a student can show me what teachers do with them inside the classroom*". A portfolio reflects teachers' performance during the school year. Hada sees the portfolio as giving " *information with evidence about their teaching and evidence about activities during school year*". Self-evaluation is an opportunity to listen to what the teacher thinks about their own performance, as Mohammed explained "*teacher can say and judge himself or herself that will be helpful for our final judgement*". While Mubarak supports the use of several tools, he has some reservations regarding the use of self-evaluation, because he believes that teachers are currently unable to use this tool. He argues that teachers need to be trained in conducting self-evaluation and that they should be provided with a standardised checklist: "*I mean a standardised form for self-evaluation in order to ensure the reliability and validity of self-evaluation*".

Inspectors are supportive of peer evaluation for formative purposes, concurring that within peer evaluation, teachers can help each other. Fahad stated "*they will have feedback from each other, which may be helpful for them*". While they support formative peer evaluation, they also argue that not all teachers have the skills to evaluate each other for summative purposes. Indeed, Hada believes that "*summative evaluation should be conducted by experts. Teachers have experience in teaching not in teacher evaluation*".

In order to support the use of a range of tools in teacher evaluation, Alia suggested that training courses and workshops should be available to teachers and evaluators (inspectors, head teachers, heads of departments) on how to use those different tools.

5.3.3 The involvement of inspectors as evaluators in teacher evaluation

This section discusses the inspectors' responses with regards their role in teacher evaluation and the difficulties that they face in executing their role. One inspector, Abdualkreem, holds discussions with individual teachers before observation, organising his timetable so he can meet teachers before and after an observation. This allows him to talk about any issues, about his visit, and gives the teacher an opportunity to ask about teaching, evaluation, content and tests, enabling relationship-building between him and the teachers. Alia, Mohammed and Hada, on the other hand, stated that they only do this when they visit a number of teachers in one school on the same day. Mohammed explained that:

> *I am very busy and I do not have enough time to talk with each teacher before their observation. I only do this if I find there are three or more teachers to evaluate in one school in my timetable. So, with three teachers I spend my day in the school so I have time to talk with them before observation*.

All inspectors, on the other hand, spoke with the teacher after the observation, discussing what the teacher did inside the classroom, providing written feedback and offering recommendations.

Inspectors also talked about some of the challenges that affect their role and the value of their role as evaluators. Firstly, inspectors see the number of teachers and the extent of their responsibilities as affecting their role as evaluators, and thus the value of their role. The number of teachers that an inspector has to evaluate varies between 70 and 260, according to subjects: social studies inspectors evaluate around 70 teachers, while maths and science inspectors evaluate around 110 teachers, while English inspectors stated that as many as 260 teachers may be evaluated by a single inspector. Because of their significant responsibilities for preparing tests, content and curriculum, meetings with department heads, supervising activities in schools and designing training programmes, it is difficult for them to visit all teachers three or four times in the school year and most of the

inspectors indicated that they often do not observe teachers for a full session, but only spend between 15-30 minutes in the classroom due to time constraints.

Second, Abdualkreem (social studies inspectors), Ali and Mohammed (both sciences inspectors) reported that some inspectors suffer from a lack of specialisation in the school levels and this affects their ability to evaluate the teachers. They stated that in high schools, chemistry, biology and physics are taught as sciences, while geography, history, sociology and philosophy are taught as social studies. The MOE chooses inspectors from social studies and sciences in high or middle and a few from primary schools. However, inspectors evaluate teachers at all school levels because there are not enough inspectors to evaluate teachers based on school level. Ali gave an example from his inspection department:

> *There is one inspector from a high school with only 8 years' experience. He was a biology teacher due to the lack of Kuwaiti teachers in this subject. He was promoted to inspector with less than 10 years' experience. He was asked to evaluate biology teachers at high schools as well as science teachers in primary schools. He found difficulties in dealing with teachers in primary schools, as he had no idea how primary teachers taught and dealt with students because he has only worked as a high school teacher*.

Third, nine inspectors mentioned the lack of training as preventing them from keeping up-to-date with innovations in the field of education and that this may affect the value of their role as evaluators. For example, Alia stated that some inspectors do not know how to use new technology in teaching, yet they evaluate other teachers on the basis of their use of technology: "*Some inspectors are unable to use an iPad, so how can they assist or help teachers in using one in the classroom? They also judge teachers for not using technology inside the classroom!*" Nawaf stated that some inspectors have been inspectors for more than 30 years, yet still depend on old pedagogies and evaluate teachers on the basis of these pedagogies and learning tools. Mubarak confirmed that many inspectors left teaching more than 10 years ago, and therefore need to be trained in current teaching practices and evaluation in order to increase the validity of their own evaluations.

Fourth, the gender of an inspector and the teacher was mentioned as affecting their ability to perform their role and thus the value of their role. Mubarak explained that as an Islamic studies inspector evaluating teachers who teach students the Quran, he looks at how words are pronounced as part of teaching the Quran in Arabic; this is called 'Tajweed'. Male inspectors find it difficult, Mubarak explained, when considering 'Tajweed' with a female teacher who covers her face with a veil. Yet as a head inspector he could not

> *let a female inspector only evaluate female teachers, as we have a large number of female teachers but our department only has a few female inspectors. We have only have a few female inspectors as many female teachers refuse to be inspectors. They prefer to be deputy head teachers and then head teachers*.

Fifth, Hadel pointed out that female inspectors have difficulties travelling to schools and that this may affect the number of teacher visits. In turn, this affects their role in evaluating teachers:

> *I do not have a car and I am afraid to drive to be honest, and there are some female inspectors who do not drive. The MOE does not provide us with a car and driver to visit teachers in schools, as we attend the educational district centre and then visit teachers in different schools in different cities. Therefore, I use my private driver, but, to be honest, sometimes I would like to visit a teacher but my driver is busy with other members of my family*.

5.4 Summary

Head teachers and inspectors shared view that the current system is concentrated on evaluating teachers for summative purposes, while they believe both summative and professional development are needed. The most common tool currently used is observation, however, teacher portfolios are also commonly used by head teachers. Both inspectors and head teachers would prefer to use multiple tools in teacher evaluation. Evaluators' role in the current system is significantly affected by their other responsibilities, numbers of teachers, lack of training courses for evaluators, lack of subject knowledge, and lack of specialisation in the school levels.

Chapter Six: An Alternative System based on a Risk-based Analysis Approach

6.1 Introduction

In this chapter, the principles of RBA are explained and reasons are given as to why it was chosen as the basis for an alternative system to the current teacher evaluation system in Kuwait. This is followed by an example of a country that has implemented the RBA approach in its evaluation.

Following the overall aim of this study, which is to contribute to making better use of teacher evaluation in the Kuwaiti context, this researcher outlines an alternative system for teacher evaluation and, following an investigation in situ, discusses whether it could be appropriate and workable as a means to overcome the challenges of the development plan that is in place for Kuwaiti education. According to the MOE (2013c), one of the development challenges for Kuwaiti education is to develop several aspects of the system, such as curriculum, management, and evaluation. This in turn, will enable the MOE to ensure a higher quality of education.

By including teachers, inspectors, and head teachers in the development of the teacher evaluation system that implements RBA, the extent to which an alternative system is capable of functioning effectively will be established. First, these teaching professionals were selected because they possess practical knowledge of evaluating teachers and working with students in schools. Second, they know what areas of teacher evaluation need to be developed or improved upon more than others who are involved in developing such systems, such as the MOE committees, even if they are not directly involved in teacher evaluation. Indeed, this could be an opportunity for them to give their input on teacher evaluation. Third, they can assess whether the idea being developed is likely to work and be valid. Fourth, they have contextual knowledge; therefore, they can make suggestions or point out aspects that would make the alternative system more appropriate and facilitate its application to different contexts. Accordingly, their participation was noted in, and suggested by, reports and research that encourages teachers and other evaluators to participate in developing or designing new systems. Darling-Hammond et al. (2012b) for example, argued that teachers and schools leaders should participate in developing an evaluation system in order to ensure it works effectively, reflects teacher performance, and

produces valid results. AlBustami (2014) also suggested that when designing, developing or improving teacher evaluation systems, teachers, head teachers, and supervisors should be involved since these people have a good understanding of teacher evaluation systems and regulations, so their understanding will help to guarantee confidence and evaluation sustainability before implementing a new system.

6.2 The concept of RBA

RBA is an approach based on an early analysis of data from internal evaluation (evaluating the risk of poor quality) before scheduling an eventual external evaluation (Ehren & Swanborn, 2012). According to Scheerens, Ehren, sleegers, & de Leeuw (2012, p. 43) that "The approach is risk-based, meaning that the investigation starts with a first screening on a limited number of quality domains and ends with a broader investigation when the risk analysis suggests that quality is insufficient". In practice this means that teachers considered at risk would be evaluated by inspector (external) a number of times, while a teacher who is considered at no risk of failing in their performance would be exempt from external evaluation.

RBA was chosen for two reasons; first, RBA combines internal and external evaluations, which are beneficial in the teacher evaluation system as discussed in the literature review provided in *Chapter Two*. RBA combines external and internal evaluations in a way that is different to the current system in Kuwait. As pointed out by Ambtelijke Commissie Toezicht II (2005, cited in Ehren & Honingh, 2011), external evaluators use the internal assurance system in order to arrange supervision and forms of inspection. As explained also by Wolf and Verkroost (2011), RBA first depends on an internal evaluator monitoring, improving and providing information about the quality of education to external evaluators. Then, external evaluators are responsible for supervision and making improvements to the quality of education, if there is risk perceived. By implementing RBA, this researcher's intention was to also support the freedom of internal evaluators to conduct evaluations, but without ignoring the importance of the external evaluator's role. Second, by using an RBA schedule, the external evaluator's role will be more focused on risk. As Ehren and Honingh (2011) have indicated, through RBA, the level of responsibility for the external evaluator moves from carrying out a full inspection to instead taking on a role that is more complementary to that of the internal evaluator's. Thus, with RBA, inspectors can distinguish between teachers with satisfactory

performance and teachers with weak performance, as the second group needs more guidance and support to improve than the first. Nolan and Hoover (2008) suggest that in any well-designed system of teacher evaluation, the procedures of teacher evaluation should differ between high-performing teachers and low-performing teachers in order to give them better direction, guidance and support to improve.

6.3 Example of RBA

In this section, the example of the Netherlands is discussed, demonstrating the freedom of internal evaluators which implements RBA to improve their own schools and showing how the role of external evaluators is to supervise and improve schools based on the data provided by the internal evaluators.

It should be first clarified that the Dutch RBA is used to cover not just teaching but also management, curriculum, school building and safety in order to attribute an overall grade to a school. The alternative system that was presented to the participants in this research, on the other hand, focused exclusively on teacher evaluation, since teacher evaluation is the main focus of this study. Furthermore, the Kuwaiti education system does not give schools overall grades or ranking so in order to propose a similar system based on the RBA that includes overall grades in the Kuwaiti context, a more comprehensive study that focuses on this particular topic would be needed.

RBA were introduced in the Netherlands in 2008. Before this date, inspections were carried out by the Netherland Inspectorate of Education (NIE) once every four years, whereas now, schools are evaluated by school boards that are expected to monitor and develop their schools by applying a quality assurance system. The NIE inspects schools when they are deemed at risk of failing (Ehren, n.d.).

According to the NIE (2012; 2009), the inspections begin with collection of data on a school. These data consist of three elements: signals, complaints and publicity; accountability documents; and student results (see Figure 6.1). Signals: complaints and publicity includes complaints from students, parents or teachers about the school and the quality of the school. Signals about low quality can also be picked up from public media, such as newspapers or social media, and from complaints made by organisations. The second element, accountability documents, consist of reports that are provided every year

to the NIE by school boards to inform them about the school's finances, its achievements and the quality of education provided by the school. The third element is students' academic achievements, such as the results of independent tests or institutional exams, which are collected to analyse students' learning at different levels to determine if their outcomes are as expected. With regard to providing accountability documents, schools should conduct self-evaluations (Nusche, Braun, Halasz, & Santiago, 2014) since this is a more logical approach (NIE, 2012).

After collecting three sets of data, the risk analysis is initiated by the NIE. This consists of two steps: primary detection and expert analysis. The first step consists of applying the standards and rules to determine the level of risk to the quality of education and then providing this information to inspectors. If the school is not at risk, it does *not* undergo any further investigation: this means that the school is trusted by NIE until the next annual analysis, and if the school continues to be deemed not at risk in subsequent annual analyses, it is only visited once every four years as a basic inspection (general inspection for all schools). However, if the school does carry some risk in terms of any of the three data elements, it progresses to the second step, which warrants further investigation by inspectors. They will investigate the school's risk and combine it with the organisational memory (previous record) of the NIE and public information about the school, e.g. its website. After that, if there seems to be nothing to cause concern, they will decide that the school does not require inspection and can be trusted until the next report filed in the next school year. If there are still doubts about the risk at that point, the school will be investigated by inspectors (NIE, 2012).

For schools that are deemed to have a certain level of risk, the inspectors meet with the school board to let it know about the school's problems and to see whether the board can resolve them. In most cases, inspectors decide to conduct an inspection to improve the school's quality by focusing on aspects that need improvement. Inspectors then write the inspection arrangement, which provides the school with an outline of the problems and some information about how to improve the school, along with a deadline for implementation of the arrangement. The final stage is intervention: the inspectors work to monitor the actions taken by the school to make the improvements outlined in the arrangement. If the school has failed to improve its quality, inspectors adjust the inspection arrangement. The school also receives intensive monitoring or sanctions are imposed. If

all identified school quality issues are resolved at this point, then the school is trusted and a verdict of "no risk" is given until the next risk analysis is provided (NIE, 2012).

According to the NIE (2009), five domains for the core frameworks of primary and secondary education are used to determine a school's level of risk. For primary schools, they are as follows:

1) Outcomes

- The students' outcomes are at the appropriate level.

2) Teaching and learning processes

- The curriculum encourages and prepares students for further education and for society.
- The teachers allow the students to take an appropriate amount of time to master the curriculum.
- The teachers are able to provide clear explanations, organise activities and encourage students to be interactive and involved in learning.
- The teachers adapt the curriculum, time for learning the subject matter, and teaching to take into account differences between pupils.
- The school climate should be characterised by safety and respectful interaction. For example:
    - Safety: students' and staffs' feelings of being safe in school are taken into account.
    - Respectful interaction: the parents are involved in the school through the school's encouragement in joining the school's activities.

3) Special needs provision and guidance:

- The teachers systematically monitor their students' progress, and the school uses standardised instruments to monitor students' learning and development.
- The school guides the students in order to assist them to develop their capabilities.
- Extra care is provided to students who need support.

4) Quality assurance

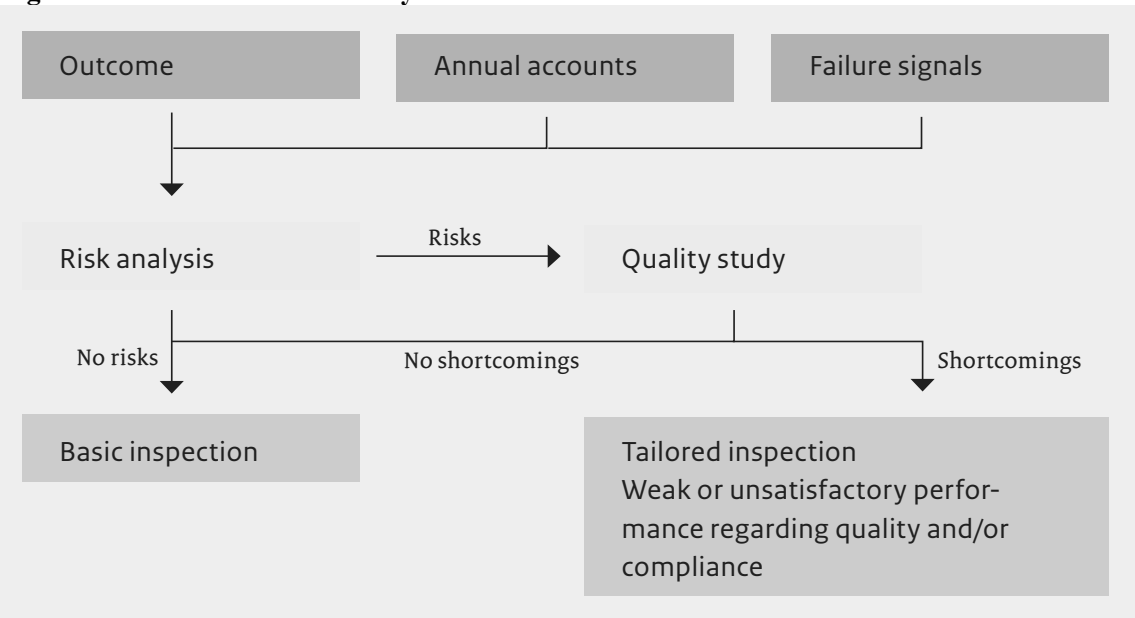
- The school has a quality assurance system such as an annual evaluation of students' achievements, regular teaching and learning evaluations, and improvement activities.

5) Statutory Regulations:

- This covers a special needs plan, examination regulations, and the planning of teaching time.

According to the NIE (2012) certain evaluation and monitoring methods are used by inspectors in schools deemed at higher risk. Firstly, students are interviewed and asked about safety, support and guidance received, teaching and attention from teachers and the time allocated to the curriculum. Parents are also interviewed to determine their involvement in the school. Teachers, the school board, and other members of staff are interviewed about the indicators of all aspects of the quality of education provided. Secondly, classes and some events or certain school facilities are observed in order to assess and evaluate the quality of education provided (NIE, 2012).

Figure 6.1: The Risk-based analysis in the Netherlands



- Cited in NIE (2009, p.5).

In terms of teacher evaluation, this falls under the responsibility of the school board. Many school boards delegate the responsibility to the school principal and the school principal may delegate to a member of the leadership team or to department heads to conduct teacher evaluations for all teachers at least once every four years. There is flexibility for the school board to design or use the framework for teacher evaluation. Overall, teachers in the Netherlands are evaluated for formative purposes in order to help them with their professional development and to provide them with support to prevent underperformance. Teachers may be evaluated for summative purposes as well, and the school board may use the results to make decision about rewards, career progression, or sanctions; however, this

use depends on the regulation of each school and school board (Nusche, Braun, Halàsz, & Santiago, 2014).

6.4 An alternative system for the Kuwaiti context

In this section, an alternative teacher evaluation system, based on the Dutch RBA, is proposed. This researcher has modified the system to narrow the focus on teacher evaluation. Based on the results of the questionnaires and the interview data reflecting the participants' perspectives, mid-year reports, final reports, supplementary documents and multiple tools were devised and included in the proposed alternative system.

Figure 6.2 shows the proposed alternative system based on an RBA approach as it was suggested to the participants in this research in order to gather their views on it and to see if they wanted to add any points of interest. This preliminary suggestion is not the final draft proposal, however, which will be discussed in Chapter 8.

6.4.1 First step: Individual teacher evaluation

This step would last from September to May every school year. All teachers have to be evaluated in this step, which consists of individual evaluation, standardised tests, and signals.

A) Individual evaluation by internal evaluators

Each teacher will be annually evaluated by internal evaluators (the head of department and the head teacher). The teachers will obtain a mid-year report on their performance and written feedback after being observed. In the summative, a final judgement will be made about the teacher's performance. The teacher evaluation ratings will be ranked from 'Outstanding', 'Very Good', 'Good', to 'Weak'. The final reports of the individual evaluation will be linked with the teacher's salary scale, annual bonus, promotions, and sanctions, such as dismissal or referral to a non-duty teaching job in the MOE.

At this step, after the evaluators make their judgements in May, teachers will be given detailed reports at the end of the school year after being signed off and accredited by the evaluators. The final report will remain private to the individual teacher, so they will be sent to the teacher's home after school ends, or can be collected from the educational district offices. The final report will consist of advice on professional development,

weaknesses and strengths, include all observation sheets written by the evaluators, and provide a detailed overall score according to the set criteria.

Teachers will be evaluated individually by using various tools of teacher evaluation, as follows:

- Observation: evaluators will observe each teacher. They have to use a standardised observation sheet in order to provide feedback to the teacher and to include this sheet in their final judgements.
- Self-evaluation: teachers will self-evaluate using a standardised checklist.
- Peer evaluation: this tool will be used for formative purposes; teachers evaluate teachers by observing each other in the classroom or examining documents such as lesson plans, assignments, and other activities.
- Teacher portfolios: evaluators will look at the teacher's portfolio in order to include different types of work done by the teacher in the school year, for example, school activities, students' assessments and progress in individual teacher evaluations.
- Student evaluation: evaluators will look at students' views about their teacher's performances. There will be flexibility for evaluators to collect this information either through interviews or surveys. Evaluators should select some classes, but not all classes, taught by a teacher (the evaluator has the freedom to choose the number); for example, teachers who teach more than five or six classes can have student evaluations from three classes.

The teacher evaluation criteria in the alternative system would be the same as those used in the current system. The criteria consist of three scales: the efficiency of individual performance, the efficiency of personality, and the efficiency of collective performance (see Appendix 1). The reason for this is that most previous studies (as mentioned in Chapter Two) on Kuwaiti teacher evaluation indicated that the teacher evaluation criteria are generally considered appropriate, with just a few needing clarification so that all evaluators follow the same interpretation.

B) Standardised tests

Schools will conduct standardised tests (between February-April, *approximately*). In May, the results of the students' results should be attached to the individual teacher's evaluation. Standardised tests should be conducted for all subjects that are taught in the national

curriculum. The school will send the results to inspection departments and they should be classified or arranged by the grades and class number; for example, Grade Four Class Three, with the name of the school, in order to facilitate the inspectors' work in the second step of the evaluation.

C) Signals

Some documents will be attached along with the final judgement given in an individual teacher evaluation. These documents are: complaints from parents, certificate of attendance at training courses and workshops (record of attending training courses), and any certifications with regard to teaching or learning, even if the teacher has obtained this from a private institution or centre for personal professional development during the school year.

6.4.2 Second step: Risk detection

This step would start at the end of the school year, in May or June (after evaluators have made their judgement and teachers have received their final report). Inspection departments for different subjects in six educational districts will receive the final individual teachers' evaluation reports, standardised tests results, and signals. The inspection departments for each subject in each educational district will arrange for a committee of inspectors to initiate the risk detection. Inspectors determine whether or not there is a risk, based on the evidence above. Risk can be detected whether there is a difference between the final reports, standardised tests results, signals (even if they are outstanding or very good), or if a teacher is evaluated as weak or good.

If the inspectors detect that there is no risk in the teacher's performance, this means that the teacher will be set on a regular evaluation (if there is also no risk in the teacher's performance in their annual risk detection). If the inspectors detect a risk in the teacher's performance, they will go through the history of that individual teacher's evaluations in the previous years. If a teacher was evaluated as 'good' or 'weak' in one year out of a certain number (e.g. five or ten), this means that inspectors will start to tailor this teacher's evaluation in the next academic year (starting in September). If the teacher is evaluated as 'very good', or 'outstanding', for several years, that means that the teacher will be set on a regular evaluation.

6.4.3 Third step: Tailored, intensive and regular evaluation

The tailored evaluation would start in the subsequent school year, beginning in September. One inspector and the head teacher should come together to evaluate a teacher whose performance for the previous school year has been detected as 'at risk'. The tailored evaluation will be the same as the first step, i.e. individual teacher evaluation as has been described, but carried out by the inspectors instead of the department head (in Section 6.4.1). Inspectors will focus on evaluating teaching and learning aspects, while the head teacher will focus on the teacher's commitment to their work and their collaboration with colleagues and school staff.

If the teacher does not improve, and the committee of inspectors finds the teacher's performance in the next annual risk detection after a tailored evaluation still at risk, the teacher will receive an intensive evaluation (the same as tailored evaluation, but conducted by two subject inspectors) for another school year. Here, two subject inspectors will evaluate the teacher, and the role of the head teacher is to evaluate the teacher's commitment to his/her work and their collaboration with colleagues and school staff. After the intensive evaluation has been conducted, if the teacher does not improve then a sanction will be imposed based on the current rules on sanctions according to the MOE on *behalf of* the CSC.

In regular evaluation, all teachers (regardless of risk) will be evaluated by the subject inspector, head of department, and head teacher. This will be done every three or four years. The evaluators focus on all aspects of teacher evaluation. This regular evaluation aims to check all parts of a teacher's performance by including both the external and internal evaluators. The regular evaluation will be the same as the first step: individual teacher evaluation as described (in Section 6.4.1), but with the participation of the inspectors. The regular evaluation will also consist of risk detection as described (in Section6.4.2)

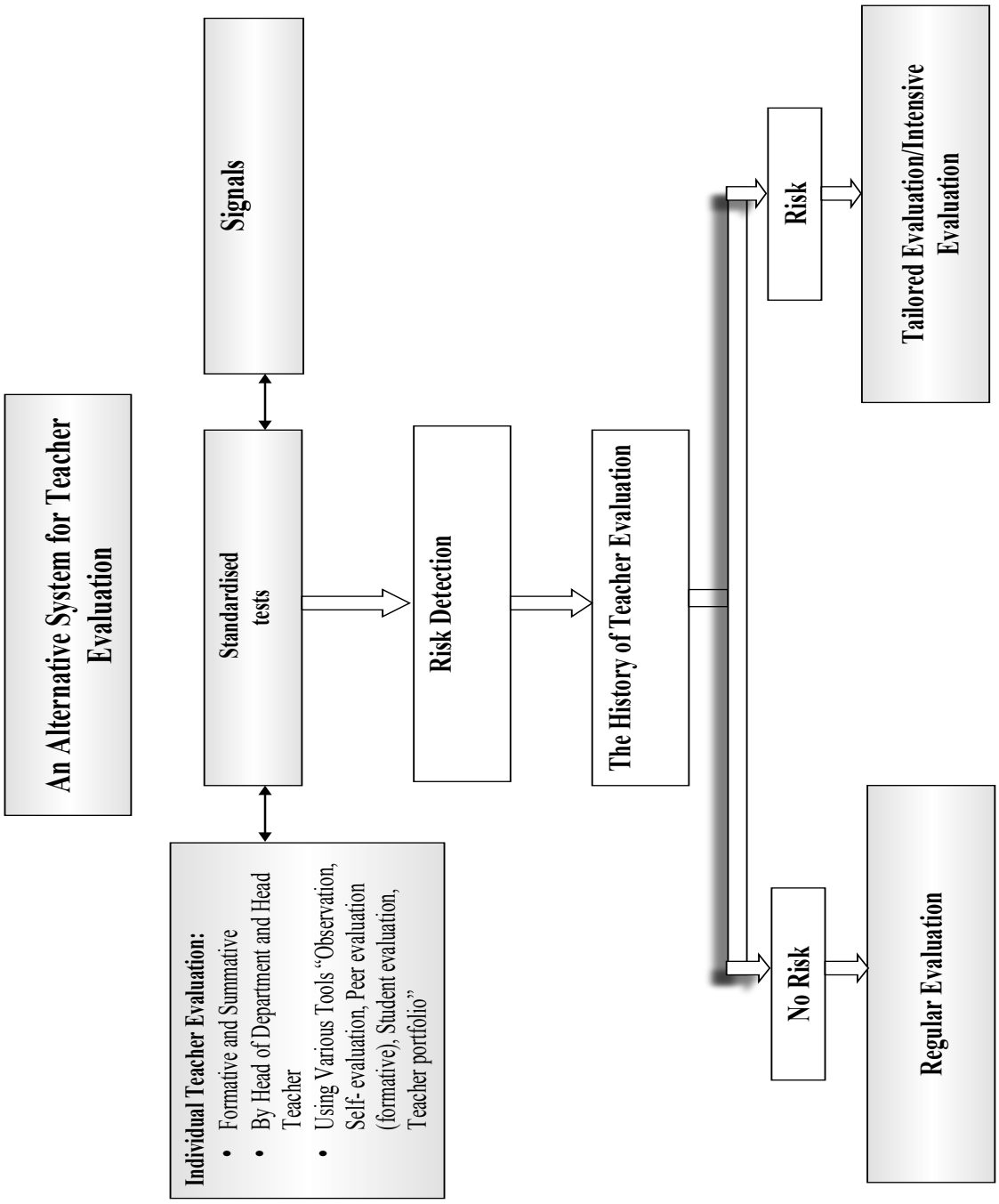


Figure 6.2: A proposal for an alternative teacher evaluation system

Chapter Seven: Participants' Views on the Proposed Alternative Teacher Evaluation System

7.1 Introduction

In this chapter, head teachers', inspectors', and teachers' comments are analysed regarding the alternative system proposed to them. As noted in *Chapter Three,* focus group interviews were conducted in nine schools, each involving five teachers (forty-four teachers, fifteen teachers in each district), while nine head teachers and twelve inspectors participated in the semi-structured interviews.

The analysis is structured according to the three main steps of the alternative system: the first step involves an individual performance evaluation, the second step involves risk detection, and the third step involves tailored, intensive and regular evaluations. The three points of views are combined for each stage, and similarities and differences discussed in each case.

On the whole the eight head teachers, twelve inspectors, and forty-two teachers responded positively to the alternative system and its appropriacy in meeting key professional development and summative aims.

On the other hand, one head teacher, Maharb suggested that the current system only needs some improvements to work well, such as providing a mid-year report and adding more appropriate teacher training by inspection departments. Three teachers also rejected the alternative system. One teacher with 21 years' experience was concerned that the proposed system would involve more work for teacher or evaluators than the current system does, and in his view, the MOE are not interested in monitoring improvements:

> *The MOE wants to make sure that I teach students and follow the national curriculum, as well as demonstrate a commitment to my work. So, evaluators visit and follow up with me to collect evidence about my teaching, commitment to work, and what I teach. The MOE does not consider whether I have improved or I need improvement.*

The other two teachers' objection to the proposed system was its inclusion of student learning outcomes, arguing that students' learning improves and develops based on both home and school experiences; if the students' parents do not care about their education, it is not fair to include their learning in the teachers' evaluations. One of them, with 7 years' experience, stated that "*it is not fair to include students' learning in my evaluation as that requires work from both teachers and parents"*. The other teacher with 34 years' experience also disagreed with the proposal to use multiple tools: *"I think observation is enough as evaluators can see from the first five minutes that I am a professional teacher"*.

7.2 The first step: Individual teacher evaluation

7.2.1 Individual evaluation by head of department and head teacher

The vast majority of participants (seven head teachers, all inspectors, and thirty-seven teachers) agreed with the proposal that all teachers ought to be evaluated by the head teachers and the heads of departments every school year, before the external evaluator is brought in. Head teachers argued that they and heads of departments are in a better position to supervise and evaluate, as they work with teachers on a daily basis, while inspectors only visit occasionally and briefly in the school year. Noriah, for example, stated "*inspectors cannot follow-up all teachers every year as heads of departments can do with all the teachers in their departments. Furthermore, we* [head teacher & department head] *know about teachers' true performance because we are with them in school"*.

Inspectors agreed with the head teachers on this issue. Hadel, for example, argued that "*I try to cover different aspects of performance, but I am not with teachers in school. So, I can cover teaching inside the classroom, and I ask heads of departments about such as a teacher's collaboration and behaviours".* Fahad also stated "*they are in-school, they know their teachers' performance more than the inspectors. Also since they are available at all times they can evaluate the teachers throughout the school year. Thereby, they can evaluate their teachers accurately*". Teachers also thought that internal evaluators are better able to conduct evaluation accurately because of their daily contact with teachers. Theoretically, inspectors can evaluate teachers but the limited amount of time they have for observation limits the value of their perspective. One teacher with 8 years' experience explained why internal evaluators are more able to get a comprehensive view than external evaluators:

The inspector had evaluated me in the previous year as weak (in inspector's grade in my evaluation). The problem was that the first time that the inspector visited me, I was at a workshop in another school, then when the inspector tried to visit me again, I was participating in a moderation meeting in the MOE. The third time that the inspector came to the school, he evaluated me and wrote my final report based on the one visit. He said that I was not committed to my work and left my class.

Another argument put forward by an inspector, Mubarak, is that giving internal evaluators the responsibility for evaluation before risk detection would allow inspectors to focus on their other responsibilities: "*Conducting internal individual evaluations for all teachers will be helpful for inspectors as inspectors will focus on other responsibilities, such as preparing tests, curriculum, follow-up with heads of departments in schools, etc.*"

One head teacher, Hasah, thought inspectors *should* participate in the evaluation of all teachers before risk detection. She argued that inspectors have experience in teaching and in evaluating teachers both for assessing and promoting professional development and that this experience benefits the teachers. Five of the teachers mentioned the neutrality of inspectors as external evaluators and their role in maintaining a balance between evaluators, leading to a more accurate judgement. One of them with 13 years' experience stated *"an external is necessary to participate as a neutral evaluator, when internal evaluators may give me an unfair judgement if I have a personal problem with one of them - especially a head of department"*.

On the other hand, many teachers did not think inspectors would be neutral since in making their evaluation they rely mainly on the internal evaluators. One teacher with 14 years' experience who is also head of department, stated that "*The inspector asks me who is good, very good, and outstanding as inspectors have a lot of teachers and overload responsibilities, and they cannot memorize the names of the teachers"*. Another teacher with 29 years' experience saw the inspector as having a very limited rather than a neutral perspective: "*Regarding inspectors that should participate as neutral evaluators. Many teachers can delude the inspector by doing a great job when inspectors visit them three or four times. They just work hard when the inspector visits them*".

With regard to Hasah's argument about the inspectors' experience in teacher evaluation, one inspector, Mubarak, suggested that inspector could have a role as guide, providing advice to heads of departments with less experience in evaluating teachers, particularly those in their first year as head of department. Wafa, inspector, also claimed *"We should not leave a head of department without support to achieve the aims of teacher evaluation"*.

Mariam, head teacher, also made an interesting point regarding head teachers being allowed to ask for help from assistant head teachers. She suggested they could play a valuable part in supporting head teachers in their role:

> *Head teachers should be allowed to ask for help from assistant head teachers when evaluating teachers as sometimes I am very busy and need some help. Assistant head teachers have the ability to evaluate teachers as they were once heads of departments. There could also be training for them in evaluating teachers before becoming head teachers themselves.*

The alternative system proposed providing a mid-year report and this was something that all the respondents thought would be useful. Asked why they supported the idea of a mid-year report, respondents thought a mid-year report encourages and motivates teachers to work towards improving themselves early in the school year (second semester). Shafah, a head teacher, saw its potential to motivate both good and weaker teachers: "*Teachers with good and weak performances at the time of the mid-year report will be motivated to improve to avoid achieving this same result when it comes to their final judgement"*. Hada, an inspector recalled the use of mid-year reports in the previous system "*Teachers obtained mid-year reports before the current system, it was very useful for all teachers in the second semester. I do not understand why the MOE stopped providing this report to teachers",* adding that "*I support providing it to teachers even though it only assists the teacher in improving by about 5%*".

Teachers generally welcomed the use of mid-year reports as they felt that receiving reports about their performance in the early months of the school year helped them to improve their performance. Some teachers also thought the mid-year report might lead to more opportunities for them to attend courses, whether provided by the Kuwaiti teacher society,

private institutions, or inspection departments. One teacher with 7 years experience explained that:

> *I receive a lot of invitations for courses in the MOE or Kuwait teacher society. If I obtain a mid-year report that shows me my weaknesses and strengths, I might think to accept invitation for appropriate courses to overcome weaknesses and develop strengths*

One head teacher, Shafah, also pointed out that mid-year reports are also a useful resource for internal evaluators, helping them to focus on those teachers whose performance has been evaluated as weak and to provide appropriate support, for example, by allowing them to select some courses by themselves or nominating them to take training courses.

With regard to written feedback after observation, all of the participants saw it also as useful. Head teachers thought feedback helps teachers to change their approaches and make immediate improvements. For example, Awatf pointed out "*Many mistakes in the classroom can be prevented from occurring by the next session through providing feedback*". Fahad, an inspector with twelve years' experience saw feedback as "*supportive of a teacher's self-reflection and improvement after observation, if the teacher has the intention to improve*". In one of the focus groups, teachers discussed whether reinstating the mid-year report would replace the need for written feedback after observation. The general consensus was that both would be necessary, as a teacher with 29 years' experience explained: "*feedback is helpful, but we miss the mid-year report and its benefits. So, we talk about the mid-year report more than feedback. In my opinion, I would prefer that even with a new system, teachers should have both"*.

In terms of providing a final report, seven head teachers and twelve inspectors thought a detailed report should be provided while Loui (head teacher) thought only the comments should be visible while the rating scores should be kept hidden. These views are discussed in *Chapter five* in the context of participants' views about what the purpose of teacher evaluation system should be. Teachers, regardless of experience, also shared the view that a detailed final report should be provided, since they would feel recognised for what they have achieved and be motivated in the next school year to improve what they have not achieved and develop their performance. For example, a teacher with 13 years' experience

stated "*final reports would make a difference in learning. I will know what exactly need to improve in the next year or know what is the aspect of my performance that is very good to develop it to be outstanding*", while a relatively inexperienced teacher stated:

> *I am a new teacher, I do not know what constitutes effective teaching or to be more clear I do not know all aspects of performance that are evaluated in teacher evaluation. Give me my final report so I know what needs to improve or be developed, or give me my reports to be satisfied with good or bad scores because I will know my mistakes (if there are mistakes) and I will not feel like I have been wronged.*

One focus group looked for reasons why the final report should not be made available to teachers in all its details. This researcher suggested to them that providing the final report to teachers may cause problem for evaluators. The response from one teacher with 10 years' experience was that the details could in any case be obtained one way or another:

> *Many teachers and I know our final individual performance score in an illegal way through some friends in educational districts, or some evaluators tell teachers the number. If I do not accept the score, I will appeal to the educational observer office (that is all, no need to make a problem if I have the right to appeal).*

Another teacher with 8 years' experience felt strongly that *"if providing the final report in detail for teachers would make a problem for evaluators with some teachers, that is not a reason to prevent all teachers in the MOE from obtaining their reports*".

One aspect that all respondents agreed on was that the final report should remain private to the individual teacher, since it contains personal information; consequently, it was agreed that reports should be sent to the teacher's home or collected from the educational district centre after being signed off and accredited by the evaluators at the end of the school year, as the researcher suggested (in the alternative system booklet).

For example, Shafah stated *"The final individual teacher evaluation reports consist of sensitive information and I believe many teachers would want to keep it private"*. Noriah

added, *"The final report is for the teacher. The teacher should have the freedom to keep it private or to discuss with his/her colleagues"*. Inspectors, such as Fahad agreed that "*Teacher evaluation reports include personal information so teachers should decide to let others know or stop them asking embarrassing questions as some teachers do not like other teachers knowing about their performance*". Hada pointed out that teachers often know who is outstanding and who is weak but that *"the details of their final reports should be kept private and teachers should have the choice of showing other the details or not*". A teacher with 18 years' experience thought that keeping reports confidential to the individual teacher would be a good idea since "*there is no need to know the reports of other colleagues. The final report is for the individual's performance*". Another teacher with 10 years experience thought that making the reports available for anyone to see would cause problems for teachers "*some teachers are jealous, knowing my score or other teachers' scores* [if the final reports distribute in school] *could make problems for us (e.g. asking head of department why this teachers and I am not)"*. Teachers were asked whether the proposed way of delivering the final reports to individual teachers would prevent comparisons among teachers regarding performance reports. The unanimous view was that colleagues could discuss their performance and help one another without the need to know the details. One teacher with 9 years experience stated that

> *If I want to ask for help or hold a discussion about performance (e.g. how to improve or what to do) I will raise the problem or concern straight away, I do not need show the teacher the details of all aspects of my performance or my scores. Equally, my colleagues do not need the details of all aspects of my performance to answer my question about a particular point*.

Hasah, head teacher, agreed that sending the final report to the teacher's home or it being collected by the individual teacher from the district office would stop teachers from negatively comparing themselves to one another; but she saw another advantage to giving teachers their final report after being signed off and accredited by evaluators, as she explained, it would stop teachers putting pressure on evaluators to change the score. Hasah was not the only evaluator with experience of being pressured by teachers to change a final report, as Abdualkreem's comment illustrates "*if teachers know their reports before accrediting, some teachers step up pressures on evaluators to change the reports*".

Participants also generally supported the linking of the final report to promotions and salary increases, as this is what already happens with the existing system, and thought it an appropriate purpose of teacher evaluation. They agree that outstanding teachers should be nominated for promotion and be awarded a salary increase. While, one teacher with 10 years' experience thought not just outstanding teachers but very good teachers should also be given recognition through promotions and salary increases: "*very good teachers deserve to become head of department and be moved to a higher salary scale, since they do not perform less well than expected"*

Participants were also asked for their opinion about the tools that would be used to evaluate them in the first step: observation, student evaluation, self-evaluation using standardised checklist, peer evaluation for formative purposes, and teacher portfolios.

Eight head teachers and ten of twelve inspectors supported using these tools in teacher evaluation, except for two head teachers of eight who did not want to use student evaluation. Their responses echoed those given in the context of the current system (See Chapter Five) regarding the ways in which how the various tools would contribute to more effective evaluations of teacher performance. Two of the inspectors, Noor and Abdualkreem, who had not been in favour of using multiple tools when interviewed about the current system, were supportive of them being used by internal evaluators in the proposed system. They argued that inspectors could evaluate teachers simply based on observation and Abdualkreem objected to the use of student evaluation, arguing that *"students are unable to participate in teacher evaluation and their emotions will affect their evaluation*".

The teachers strongly supported the use of multiple tools in the first step of the alternative system. In their view, there are advantages to using multiple tools: firstly, their performance will be evaluated more accurately as evaluators will get a clearer picture of their performance which will not depend on observation alone at a particular time; this would contribute to more reliable mid-year and final year reports that accurately reflect their performance, leading some teachers to work harder (such as those teachers who do a great job when evaluators observe them only inside the classroom); the inclusion of teacher self-evaluation, students, and their work (portfolio) in the information evaluators are given about individual teachers would reduce subjectivity. The strength of their feelings about

multiple tools can be glimpsed in the following statement made by five teachers in one of the focus groups: *"using various tools of teacher evaluation is like a dream"*.

Teachers also thought that the best way for teachers to share their experience is by visiting each other in the classroom and providing feedback to each other with a formative purpose. Teachers with 9 years experience, for example, stated "*I don't want my colleagues to evaluate me and use their opinions about my performance to make decisions about me. What would be good is if my colleague who observes me helps me rather than judges me*". Three teachers were not in favour of conducting peer evaluation, even for formative purposes, believing that in the absence of a consensus among the teachers, peer evaluation would result in subjective judgements. Since evaluators may take into account what a teacher says about his/her colleagues, this could indirectly affect their judgement about a teacher's performance.

Views about the use of students' evaluation varied considerably among teachers. Six teachers objected to the use of students' evaluation, arguing that students have no idea what teaching is and so are not able to determine the quality of teaching. They also pointed out that students do not like strict teachers, so even if a teacher is outstanding, if they are perceived as strict by the students, they will be evaluated negatively. Others teachers supported the idea of using students' evaluation, arguing that since it is the students who are interacting with teachers every day, their views should be taken into account and that they could contribute valuable information about a teachers' performance. Whilst agreeing that students have no idea what teaching is, some pointed out that students could be asked questions that would help to determine indirectly how they experienced the teaching. For example, one teacher with 8 years' experience stated:

> *Students have to participate in reflecting on performance, as they are the most important element in the teaching and learning process. They are with the teachers in class and can give an impression to evaluators about our behaviour inside the classroom, whether we respect students, and take care of them and their learning and more*.

Participants were asked about their reasons for supporting the use of a standardised checklist for self-evaluation. Many of them see a standardised checklist as supporting the

teacher in making judgements about themselves, while some others see a standardised checklist as ensuring that all aspects of performance are addressed in the self-evaluation. Waleed, a head teacher, stated that *"Some teachers face difficulties when they talk about themselves. A standardised checklist supports teachers in making judgements"*. Alia, an inspector, also stated "*Using a standardised checklist will make self-evaluation easy. Teachers will be asked particular questions that reflect how that teacher performs"*. A teacher with 13 years experience saw the standardised checklist as making self-evaluation easy to conduct, as "*I can answer questions better than take time to think about my performance then evaluate myself"*. While Mariam, a head teacher, stated "*to make self-evaluation cover all aspects, as some teachers will give judgement about some aspects and may forget to cover other aspects*" and similarly, inspector Nawaf saw the standardised sheet as ensuring that teachers *"don't forget something important about their performance*". A teacher with 7 years' experience stated "*Direct questions about my performance will make my judgement more accurate when conducting self-evaluation. Since I do not know what to include to reflect my performance during the school year*".

Mohammed saw the benefits of the standardised checklist very much from the evaluator's perspective: it makes teachers' self-evaluation easy to analyse and evaluators can easily compare it to the criteria of teacher evaluation: "*It is also good to compare teachers with his/her colleagues in the same school and other schools to help me know he is a good teacher"*. However, one teacher with 2 years' experience rejected the use of the standardised checklist, arguing that it would, "*reduce self-evaluation",* stating a preference for *"writing a short report about my performance*".

7.2.2 Standardised tests

All head teachers, all inspectors, and most teachers support the use of standardised tests and attaching the results of the students' scores to the final individual teacher's evaluation reports as evidence to be used in the next step, as suggested by this researcher.

Participants thought that students' scores could be used as evidence of a teacher's effectiveness in aiding their students' learning. For example, one head teacher, Loui, stated "*standardised tests are very useful as the quality of learning can be aligned with the quality of teaching to determine the contribution of teachers to students' learning"*. One inspector, Hadel thought that teacher evaluation should not focus only on a teacher's

performance*" but also evidence of students' learning as obtained by standardised tests"*. One teacher with 13 year's experience thought that including results from standardised tests would ensure that "*the quality of learning will be covered as well as the quality of teaching. There is a relative relationship between them, which determines a teacher's efforts*".

One head teacher, Noriah, thought that including these results would also motivate teachers to care about improving students' learning:

> *The focus in teacher evaluation will be both on teaching and learning, more than on non-teaching duties, which means that teachers who have not improved students' learning will be asked to leave teaching as they cannot handle the demands of the job*.

Many teachers also thought standardised tests might serve as a shield for teachers, using them to make appeals if they felt their final evaluation report was unfair. For example, a teacher with 29 years' experience stated:

> *Students' results from standardised tests can be used to support my appeal if I do not accept the evaluators' judgement. However, within the current system, if I appeal there is nothing that will be changed because the educational observer only reviews the observation checklists and the final evaluation report.*

In contrast, one teacher with 13 years experience from Asimah district did not favour standardised tests being used, arguing that the standardised test does not distinguish between outstanding teachers and inadequate students or vice versa and does not account for the influence of family, curriculum and school environment on students' level of achievement.

Even teachers who were supportive of standardised tests being used in evaluating teacher performance had a number of concerns. First, if teachers are on leave or move to another school in the middle of the school year, and another teacher takes his or her place, it would not be possible to attach standardised test results to the report of the replacement teacher. Second, account has to be taken of students' starting point. As three of the teachers pointed

out, some schools put students with weak academic records in one class and ask teachers with a long history of outstanding teaching to teach them. Third, students' weakness in the Arabic language affects students' results in other subjects, e.g. Islamic studies, social studies, and science, though, one teacher with 18 years' experience suggested a solution to this in her own subject:

> *Arabic language is the main problem for many students in primary school. Therefore, the standardised test for my subject (science) should include images and questions such as making the link between the question and the answer, and true and false. I mean reducing questions that ask students to write and require them to read carefully, since the test is not a language test*.

Concerns were also raised about how inspectors should consider students who have dyslexia and dyscalculia (special educational needs) as their schools are integrated (the MOE applies integration in some schools in different educational districts). One head teacher, Noriah, took a different perspective, expressing concern that high marks might not always accurately reflect student learning:

> *Inspectors should consider the risks of high marks as well as low marks. Some teachers with high marks may teach their students techniques to achieve these marks on the tests and thus teachers may not actually be good at contributing to their students' learning.*

With regard to who should conduct the standardised tests, this researcher suggested they be conducted by schools and the results sent to inspection departments in each district. However, there was concern that schools did not have the necessary expertise to conduct the tests. Ghadeer suggested that:

> *I think that standardised tests should be set by experts for accurate results. MOE's centre for evaluation and measurement (National Centre for Educational Development,* [NCED]*) can conduct these tests, from design to marking and sending the results to inspection departments. As NCED has staff that are experts for standardised test.*

Two head teachers felt that although school staff are not experts, the school should still be involved in the process, as Noriah suggested: *"the centre can ask the school to help them to administering the tests in order to make it easier to conduct (not analyse the results) as this centre does not have a large staff*". Waleed elaborated on this potential collaboration between the NCED and individual teachers, suggesting that:

> *Schools can participate in administering and mark the tests, as every teacher can observe the test for a different subject and also ask physical education teachers and art teachers for their help. With regard to who marks the test, every department can mark their own test by covering the students' names. Then, the results can be sent to NCED to be analysed and sent to the inspection department.*

Teachers thought it should be the NCED in the MOE that designs and analyses the standardised tests as an expert party. For example, one teacher with 7 years experience stated "… *centre staff have more expertise ... they know how to design and mark standardised tests*". The vast majority of teachers support being involved in administering standardised tests in school and then sending them back to the centre for analysis. One teacher with 12 years experiences stated "*this is a possible way to administering tests since the NCED does not have a large staff to administer tests in schools*". Another teacher with 7 years experiences commented "*we administer tests from this centre every school year, for the fifth stage in primary school. Standardised tests are impossible to administer in all schools in Kuwait without our help*".

All the inspectors agreed with Ghadeer that since a centre already exists within the MOE (NCED) with the necessary expertise, the standardised tests should be designed and marked by this centre. Some of them, e.g. Mohammed and Nawaf, shared the same opinion with head teachers and teachers that inspectors are unable to create and analyse the standardised test as well as the experts in this centre can do**.** However, other inspectors expressed a preference for the tests to be designed and analysed by a neutral party. For example, Noor explained that the *"inspection department should not be involved in teacher evaluation before risk detection"*. She also disagreed with those who stated that inspection departments do not have experience in conducting standardised tests: *"Inspectors are able*

to carry out standardised tests since in our department, [Arabic inspection department] *standardised test are conducted every school year"*.

7.2.3 Signals

All head teachers, inspectors, and teachers agreed that, in addition to the standardised test results, certain additional documents should be attached to the final individual teacher evaluation report, as this researcher suggests in the booklet. These documents would be used as signals and would consist of reports about certificates of attendance for training courses and workshops, including any certifications that teachers have obtained in the school year with regard to teaching or learning even if obtained from private institutions or centres as part of their personal professional development.

Head teachers, inspectors and teachers agreed that the inclusion of these documents in the final evaluation report would encourage teachers to participate in training courses and workshops, and more generally, to engage in personal and professional development, not just the weaker teachers but the good ones too, as Shafah, a head teacher, explained: "*Documenting training courses and workshops attended within the final teacher evaluation report will motivate teachers to attend courses put on by inspection departments or other schools– especially very good teachers and outstanding teachers*". Waleed thought this aspect of the alternative system is important in that it would acknowledge those teachers who engage in development activities: "*this is a great point, to support personal professional development, as well as protect the right and effort of teachers who work on themselves by attending courses*". From the inspector's point of view, Hadel saw the inclusion of these documents as supporting and motivating teachers *"to be involved in professional development, and other such activities to obtain certificates that can be attached to their evaluation*". For Mohammed, another inspector, these documents would help him to see what teachers had done over the school year to develop their teaching: "*there will be a record of training courses in signals that can show us the difference between teachers who work to develop themselves and other teachers who do not, which will motivate them to get involved in training courses*". From the teacher's view, one teacher with 9 years' experience spoke of the desire to have their efforts acknowledged:

If my personal development or attendance at courses in the MOE will be recorded, I will be motivated to involve myself in professional development, as my efforts will be acknowledged and distinguish between me and other teachers who do not work on themselves.

Ghadeer also suggested adding a report about a teacher's attendance and absences, and warning letters for late attendance**,** as being valuable evidence of teachers' commitment to their work. Inspectors also saw the value of such documents in reflecting other aspects of teacher performance outside of the classroom. Ali, for example, thinks "*these documents will give a picture of a teacher's commitment*". Teachers agreed with inspectors and head teachers regarding the value of including this kind of documentation, and in addition saw these documents as potentially useful when making appeals regarding their final teacher evaluation reports. In one of the focus groups, the following comment by one teacher with 7 years' experience gained the unanimous support of the other members of the group "*We do that, we commit to work. So, there is no fear of including some documents on what we actually do as these documents will not affect our score but may support our objection and show the truth*".

There was less consensus regarding the inclusion of parents' complaints, however. Four head teachers, one inspector, and seven teachers agreed that the signals should include complaints from parents, in that this would be a way of involving parents in teacher evaluation. For example, Mariam, head teacher, stated "*teacher evaluation criteria include dealing with parent. Complaints will show the way teachers deal with parents*". Another head teacher, Hasah stated "*it could be useful to include parents' views in teacher evaluation, especially with weaker teachers, as it will help to confirm our judgement*". Furthermore, Loui, a head teacher, thought that including parent evaluation would have a positive effect, in that "*This could make teachers care about parents, as some teachers do not care about parents' complaints. I have tried to solve similar problems many times but some teachers do not avoid creating the same problem in the future*". One inspector, Hadel, saw parental complaints as making an important contribution to risk detection:

If there are complaints about the performance and behaviour of an individual teacher, these should be included as they are helpful in making a judgement in risk analysis. An evaluator should look at a teacher's performance and their behaviour as teachers in teacher evaluation.

Seven teachers agreed with including parents' complaints as an indication of issues that may be occurring for their child. For example, one teacher (10 years' experience) stated:

Parents follow up and teach their children at home, so they are a part of learning. They can add some important information about teachers (e.g. mistakes in tasks by teachers or wrong things, such as in Mathematics or English when they 'teach students wrong') that may be uncovered by evaluators.

A teacher with 11 years' experience also thought including parents' complaints was a useful safeguard against bad teaching:

Some teachers behave badly in dealing with students but many students will not tell a counsellor in school or a head teacher or the head of the department. Instead, they state the problem to their parents and parents state the problem to a counsellor or head teacher. By including parents' complaints about such bad behaviour, this kind of problem might be mitigated. Those teachers will respect students if their parents can complain and their complaints will be attached to their evaluation

In contrast, four head teachers preferred not to include complaints from parents, arguing that many complaints from parents are malicious and false, as students may say things to their parents about teachers that are not true. Eleven of twelve inspectors thought that complaints from parents would not be useful in evaluating teacher performance, especially when determining whether the teacher needs tailored evaluation or not. For example, Abdualkreem, felt that such complaints should not be made public and should instead be kept within the school and be addressed by the school's management: "*We cannot judge a teacher on the extent to which parents are happy with him/her. The complaint will be resolved when school management addresses it or the MOE investigates the complaint by other departments*".

Most teachers also rejected the idea of including parents' complaints. Some of them pointed out that parents only intervene if students fail in particular subjects, in which case they come to school to complain about those teachers. Some of them argued that parents' complaints are resolved in school by the school's management, so there is no need to include complaints that are already resolved. Others pointed out that parents may overstate matters and indeed, that some parents may exaggerate praise for a teacher as a way to ensure that the teacher will help their child. Conversely, parents can have unrealistic expectations about the amount of individual care a teacher can have for their child, putting a teacher in an impossible position, as this female teacher in a boys' school recounts:

> *One of my students asked me to go to the toilet and as he was running back to the class, he hurt his head. The following day, his father came to make a complaint about me saying that I did not take care of his son. The father said that I should have taken his son to the toilet and waited for him due to his son still being a child* (*he was in Year 4)* [students are roughly nine years old]*, but to do what his father said, I would have had to leave my class of 25 students to take care of his son*.

7.3 Second step: Risk detection

All participants agree on the inspectors' role regarding teacher evaluation, which starts after the final individual performance reports, students' scores in standardised tests, and signals to detect a risk with teachers' performances have been gathered, as suggested by this researcher in the booklet.

Inspectors shared the view that although they would not be directly involved in teacher evaluation, their role would be to regularly check teachers' performance as a way of monitoring the teachers and quality of their teaching. For example, Mubarak stated "*we will not be involved in evaluating all teachers, but our eyes will be open on all teachers*" and Fahad added "*we are closed to teachers*". Head teachers' agreed that the inspectors' role should be to monitor teachers' performance and to detect risk based on the internal evaluators' reports and other evidence. Loui, for example, thinks that *"through this step, the inspector can monitor teacher evaluation, instead of participating in teacher evaluation for all teachers as an external evaluator"*.

Furthermore, inspectors and head teachers agreed that the evaluation report, students' score, and other evidence, would help inspectors in analysing teachers' performance and detecting risk. As Ali, an inspector stated, "…*with this evidence, I can make decisions about a teacher's performance if there is a risk with a teacher's performance*". Shafah, head teacher, thought the proposed array of documentation would facilitate risk detection*, "as inspectors through individual evaluation report will analyse the teaching practice, standardised tests will analyse the quality of learning, and other documents will analyse different things related to a teacher's performance"*. Teachers also agreed that the inspector's role would be to check and analyse different parts of their performance based on different sources of evidence. For example, one teacher with 18 years' experience stated: "*inspectors will analyse a teacher's performance not only based on what evaluators said or have seen inside the classroom, but the evaluator will look at learning outcomes, and professionalism*".

Inspectors, head teachers, and teachers also thought that discrepancy between the final teacher evaluation reports and other evidence could signal risk, or if teachers are evaluated as weak and good performance, as per the researcher's suggestion in the booklet. Mohammed, inspector, thought that *"weak and good teachers pose a risk based on their score, but inspectors also should have freedom to determine the risk with outstanding and very good teacher based on standardised tests and signal*s". Noriah, head teacher, also thought that *"inspector should detect risk if there is no link between outstanding or very goof performance reports, and other information. This kind of checking is needed to see if teachers are really outstanding*". Another head teacher, Mariam also thought inspectors should participate in evaluating those teachers who are good "*a good performance may become weak the following year if a teacher does not receive sufficient supervision. So, the inspector can focus on both weak and good teachers in order to prevent deterioration in the performance of good teachers*". Teachers shared the same view with other participants. For example, as one teacher with 8 years' experience put it "*Inspection departments should have the freedom to determine risk even though the score of the teacher evaluation is outstanding, since inspectors have other standardised results, and signals. They can use to see if a teacher is really outstanding*". Another teacher with 10 years' experience thought

> *I do not like teachers to obtain outstanding score, if they do not deserve this score. By allowing inspectors to detect risk if there is no match between the results of the three elements* [the final individual evaluation report, standardised test, signal] *that will expose fake outstanding teachers (weak teachers who do not work hard and obtain an outstanding score, due to a personal relationship or for other reasons). That will make me feel less depressed.*

The effect of personal relationships on teacher evaluation and in particular, on the final report, was also a concern for Noriah, a teacher with 10 years' experience "*linking the standardised tests and signals with the final evaluator report will reduce the effect of personal relationships on teacher evaluation*". She explained teachers who have a good personal relationship with head teachers, will be less able to put pressure on head teachers to give them a high score, as inspectors will be able to tell from the other sources of evidence whether a teacher does deserve to be evaluated as outstanding.

Participants also agreed that in the risk detection step, a committee of inspectors should make decisions about risks regarding an individual teacher's performance, as suggested in the booklet. The head teachers and inspectors thought a group of inspectors would be more effective than one inspector in making decisions about risk because, as head teacher Awatf explained "*a group of people means different opinions and perspectives will be presented that will lead to a valid decision*". Hada, an inspector, stated that:

> A *teacher may be identified as posing a risk based solely on a subjective opinion and, this is more likely to happen when one inspector conducts the risk detection, whereas with a group of inspectors, subjectivity can be reduced as one inspector cannot impose their opinion without evidence*.

Teachers also thought a committee of inspectors should implement this second step for similar reasons: a group of inspectors would lead to more accurate decisions about risk, since perceptions about risk might differ from inspector to inspector. Teachers, as other participants, thought that a committee's decisions was less likely to be influenced by personal relationships since some teachers may have personal problems with particular inspectors, thereby reducing the element of subjectivity.

All participants (except for one head teacher and five teachers) thought that if no risk was detected, teachers should be trusted and receive regular evaluation, unless inspectors found risk in the next annual risk detection, with individual teacher evaluation by internal every school year. Head teachers shared the belief that trusting outstanding and very good teachers is a kind of motivation for them to keep up their standard of teaching and indeed, improve upon it, as Awatf commented: "*Teachers identified as not at risk will be motivated towards further development due to the fact that they have been trusted by a risk analysis process*". Inspectors concur, with Fahad stating "*outstanding teachers do not need to be evaluated externally every year. Let them feel that they are trusted and supported to do a great job every year*".

Teachers had varying views on how outstanding and very good teachers needed to be evaluated. A teacher with 15 years' experience stated "*if I am outstanding teacher and there is no risk in my performance, annual risk detection is enough. Inspectors will not add something for me*". Other teachers confirmed the comments made by evaluators, in stating that feeling trusted would encourage them to keep working hard. For example, a teacher with 7 years' experience said "*I like this idea, it makes me feel I am really outstanding ... I will keep doing my best*". Another teacher, with 9 years' experience also thought that feeling trusted would encourage her to be very good or outstanding every school year to avoid the loss of reputation that would occur if she were to be identified as at risk:

> *I will do my best to be trusted and to avoid the disgrace of being detected as at risk in my performance, which would result in loss of reputation or respect. Especially when my colleagues see the inspector visiting me in the classroom, while they are evaluated by the internal evaluator because they are very good or outstanding*.

Participants also agreed with the suggestion that once a risk has been identified in a teacher's performance, the teacher should receive a tailored evaluation. However, inspectors should look at the history of the teacher's performance before making the decision to evaluate them directly. If a teacher has been evaluated as outstanding in previous years, inspectors can choose to trust them even if they have currently been identified as at risk.

With regards to the importance of looking at a teacher's history, Ghadeer, head teacher, stated "*teachers should not be put in a circle of risk because of one year in which they performed badly*". Another had teacher, Loui thought looking at a teacher's past record is important because it enables evaluators to make a distinction between teachers who need tailored evaluation and those whose weak performance is temporary: "*Some teachers may be outstanding teachers but in a particular year they may have problems or circumstances that affect their performance*".

All inspectors agreed that being able to refer back to previous evaluations would help them in knowing when it was appropriate to give teachers another opportunity to improve by themselves without intervention, as Mubarak explained:

> *If the teacher had an outstanding performance for five or ten years, then it does not make sense to identify them as at risk just because they've had one bad year out of the past 10. Absolutely, there is something wrong; so as an inspection team we should give the teacher another opportunity to improve by themselves within a year without intervention*.

Indeed, Alia argued that "*if there are teachers with outstanding outcomes for at least the last five years, they can improve themselves with internal and training courses or workshops*".

In this regard, teachers too thought it important to take a teacher's past performance record into consideration, since teachers may suffer from problems that may affect their performance in a given year. Therefore, sometimes it is appropriate to give them the opportunity to improve by themselves during the following academic year. Female teachers were strongly in favour of this approach since as one female teacher with 6 years' experience explained:

> *Many female teachers have taken maternity leave and they go back to school in the last three months of the school year. They do not obtain outstanding or very good because they have returned to teaching at the end of the school year so evaluators do not judge them as outstanding even though they deserve to be. With alternative system, if teachers have previously obtained outstanding, but*

then don't because of maternity leave, they will be given the opportunity to do so again in the next year before being given tailored evaluation, as they may not need it.

The five teachers who thought inspectors should be involved in the first step, though that teachers at risk should receive intensive evaluation conducted by two inspectors. They rejected the idea of tailored and regular evaluation as proposed but did agree that a teacher's history should be taken into account before making the decision to conduct an intensive evaluation, in order to give teachers opportunity to improve. Hasah, a head teacher, also thought that the inspector should participate in the first step and that intensive evaluation should be conducted by two inspectors but rejected the idea of looking at a teacher's history before making decision about intensive evaluation, believing "*if there is risk, the history does not make sense*".

Finally, one inspector, Abdualkreem thought that what would be helpful in the risk detection step would be if the inspection department were to provide a list of training opportunities:

In risk detection, we can read the needs of teachers as their reports are available to us as an inspection team. Thereby, we can identify the areas in which there is the greatest need for training among teachers and prepare a list of training courses that could be appropriate for them.

7.4 Third step: Tailored, intensive and regular evaluation

With regard to *tailored evaluation* for teachers who pose a risk, one subject inspector will replace department heads in the first step: individual teacher evaluation. Teachers who do not improve and are still at risk in the following annual risk detection will receive *intensive evaluation* within the next year but this time, two subject inspectors will be involved in the individual teacher evaluation.

Head teachers believe the benefit of inspectors' participation in evaluating teacher with risk is that they can pay more attention to those teachers. For example, Waleed stated the advantage as being that "*Inspectors will focus on teachers who pose a risk in their performance, in order to improve their practice. Inspectors will not waste their time with*

teachers who do not need inspection due to the fact that they are outstanding". Noriah concurred, adding that "*The focus of the inspector's role can be on improving teachers who need improvement*". Shafah also saw this approach as an efficient use of the inspector's expertise and their limited time: "*Inspectors will not need to spend time with all the teachers…. thus teachers who need improvement will be supported by a subject expert, such as an inspector*".

The proposed approach to dealing with teachers at risk was seen by inspectors as saving time, through judicious role allocation between internal and external evaluators. As Mubarak stated:

> *Inspectors will have a number of teachers at risk and inspectors will know what their weaknesses and risks are with regard to teacher performance before starting the tailored evaluation. Therefore, the inspectors will not waste their time determining the strengths and weaknesses, but can start from where the internal evaluators ended, by focusing on improvement.*

Teachers also believed that if inspectors only focused on some teachers that are at risk, they would also be able to do the follow up. For instance, one teacher with 18 years' experience stated *"... inspectors can also follow them up step by step as inspectors will only evaluate teachers who are at risk – not all teachers".* Another teacher with 12 years experience argued "… *so, there will be no excuse for insufficient follow up with teachers, and asking the head of department to give a report to him/her* [inspector] *about the teacher".*

Some inspectors also thought tailored or intensive evaluation would enable inspectors to address some of the problems in the current system. For example, Abdualkreem, Mohammed and Ali referred to the difficulties with evaluating the teaching of social studies and science due to a lack of specialisation at all school levels. The proposed system would resolve this as by only evaluating teachers at risk, it will be more possible to match up inspector and teacher according to level and subject, so an inspector from primary would evaluate teachers in primary schools and so on. Mubarak also points out that by applying tailored and intensive evaluations, the proposed system may also solve the gender issue as

the inspector may ensure that female inspectors are with female teachers, and male inspectors with male teachers.

Another benefit of tailored and intensive evaluation as pointed out by one teacher with 29 years' experience is that any difficulties that arise between inspector and teacher can be addressed more easily:

> *The head of an inspection department in an educational district can monitor inspectors who participate in evaluating teachers at risk. That may lead to solving problems that may arise between teachers and inspectors. Teachers can also complain to head inspectors to avoid some personal problems or other types of problems before the problem affects their performance.*

Participants also supported the idea of imposing sanctions on a teacher who does not improve after intensive intervention. For example, one teacher with 8 years' experience suggested that *"a teacher who does not improve after intensive evaluation does not deserve to be a teacher. It does not make sense to allow this teacher to teach if there is no improvement after tailored and intensive evaluation"*. Wafa, an inspector, thought that a teacher who is not able to improve after intensive and ongoing input from evaluators, should be dismissed. Head teacher, Loui expressed a similar view: "*if after internal, tailored and intensive evaluation a teacher still does not improve, the problem is the teacher, and he should not teach students anymore*".

In terms of the division of responsibility between evaluating teaching and learning on the one hand, and teacher's commitment to work and collaboration on the other, participants had a range of views. In the suggested system in both tailored and intensive evaluation, the inspector would concentrate on teacher and learning aspects, while head teachers would focus on a teacher's commitment to their work and collaboration with staff and colleagues. However, there is variance between participants' views on which aspects should be evaluated by inspectors and head teachers in tailored and intensive evaluation.

Eight of the twelve inspectors preferred to work to improve teachers' performance without the help of internal evaluators. Hadel reasoned that "*Internal evaluators do what they can with teachers before inspectors detect risks. Therefore, let us see what the inspector can do*

with teachers who pose risk". Fahad thought inspectors *"should have the freedom to find out and improve what we determine as risk".* Thirty-two teachers thought inspectors should concentrate on teaching and learning, the main rationale being that heads of departments and head teachers play their part in the first step of individual teacher evaluation before risk detection, and therefore inspectors should be the ones to identify the risks and then concentrate on improving those aspects in a teachers' performance.

By contrast, three of the head teachers thought head teachers should also focus on teaching and learning to avoid the potential subjectivity that could arise if only inspectors evaluate all aspects of performance. Mariam stated: *"... because one person's judgement is subjective. In the first step, two internal evaluators should be used, and similarly, in a tailored evaluation, two evaluators should also evaluate the teacher in teaching and learning".* Four out of twelve inspectors thought head teachers should have a role in in all aspects, agreeing with Mariam's view that multiple perspectives reduces subjectivity. For example, Noor stated that "*teachers should be evaluated by two evaluators, especially in teaching and learning aspects of teacher evaluation so as to be more accurate, and more credible*".

Other four head teachers thought that they and the inspectors should evaluate teachers in teaching and learning, as well as their commitment to their work. For example, Shafah stated "*if either internal and external evaluators participate in teacher evaluation, each evaluator should participate in the full evaluation of teaching, learning and commitment to work*". Awatf claimed "*they are a teacher in our school, I am a manager and I should therefore participate with inspectors in evaluating that teacher, I do not want to lose my power*". Ten teachers also thought that head teachers should participate in all aspects of teacher evaluation. For example, one of them with 15 years' experience stated,

> *Although both inspectors and head teachers have left teaching, they both have experience. But head teachers have more experience in the educational field than inspectors. I believe that head teachers can help teachers as well as advise inspectors on some aspects of teaching or learning.*

One of the head teacher participants, Mariam, subsequently changed her mind about the head teacher's role in intensive evaluation, and agreed that inspectors should evaluate

teaching and learning aspects, while head teachers focus on a teacher's commitment to their work and collaboration with staff and colleagues: *"two evaluators, even though they are both external, and I will focus on commitment to work. With two evaluators (inspectors), I believe subjectively will be avoided"*. The four inspectors who thought that head teachers and inspectors should focus on both the teaching learning aspects, and the commitment to work and collaboration in tailored evaluation, also changed their opinions with regards to intensive evaluation, with Noor arguing *"two evaluators can conduct evaluations for teaching and learning aspects that are more accurate and reduce subjectivity"*. Noor came around to thinking that *"head teachers are more able to focus on evaluating commitment to work than inspectors, as this is a part of school management"*. Two teachers also changed their opinions regarding the role of inspectors, with one teacher with 9 years' experience stating:

> *After intensive evaluation, sanctions may be taken against teachers, so I think two inspectors should participate so as to provide careful and accurate teacher evaluations. As head teachers are the ones who will evaluate all teachers, I don't think they would be able to conduct intensive evaluation as well*.

Another head teacher, Hasah, who thought inspectors should have a role prior to risk detection in the second step, thought that two inspectors could evaluate teachers effectively without head teachers' participation, stating "*let them take their opportunities to improve teachers*".

With regard to regular evaluation for all teachers (regardless of risk), all participants (except Hasah and five teacher who thought inspectors should participate in the first step every year) supported the participation of inspectors with internal evaluators to evaluate all teachers. They agreed that regular evaluation and the participation of external evaluators is necessary in order to ensure that the performance of those teachers evaluated as 'outstanding' by internal evaluators does not pose any risk. One inspector Alia, argued that "*internal and external evaluators should collaborate to achieve the aims of regular evaluation, which are checking and evaluating all teachers (including those not at risk)*".

Participants thought that all aspects of a teacher's performance should be evaluated by both internal and external evaluators in regular evaluation. Given the limited time inspectors

had to visit all teachers in their classrooms, compared to heads of departments and head teachers, they thought that without the input of the internal evaluators, inspectors would not be able to make accurate judgements. As one inspector, Hada commented:

> *All teachers have to be evaluated but inspectors will not be able to visit teachers in the classroom many times in a term. Therefore, internal evaluators should be involved in all aspects of teacher evaluation to make the most accurate judgements*.

Participants' views differed with regards to how often regular evaluation should be conducted and can be divided into two groups. The first group, comprising twenty-one teachers, elven inspectors, and seven head teachers, support regular evaluation every three years. Mohammed's view was shared by this group: "*inspectors should conduct regular evaluation every three school years. I believe that inspectors should not leave the outstanding or very good teacher for more than three years without at least checking in on the classroom*". And head teacher, Shafah concurred, stating that "*four years is too long, the teacher should be evaluated by external and internal evaluators for more credibility at least every three years*". The second group, which included sixteen teachers and one inspector, supported regular evaluation every four year, believing that four years is a suitable length of time to leave between evaluations of teachers who have been appraised as outstanding or very good. One teacher with 7 years' experience stated that *"regular evaluation is not about formative and summative purposes, rather it is about checking teachers with no risk. Four years is a good time"*. The inspector in this group, Mubarak, agreed with the teachers, arguing that: "*three years is not so different to four years, four is a good time to check all teachers, as during those four years so we should concentrate our attention on those teachers who need to improve".*

Finally, Mubarak pointed out an interesting point about teachers who are given intensive evaluation at the same time as conducting regular evaluation, suggesting that *"two subject inspectors should be free to evaluate teachers in intensive evaluation in each district, while all inspectors participate in regular evaluation*".

7.5 Implementing the alternative system: some considerations

The analysis of participants' views has highlighted a number of points that should be taken into account when implementing the alternative system:

- Training and workshops should be provided for teachers to introduce them to self-evaluation and peer evaluation, as well as to preparing a teachers' portfolio.
- The number of inspectors needs to be increased so that inspectors can focus on teachers alongside their other responsibilities.
- Training and workshops for all evaluators (internal and external) are needed on how to conduct the alternative system. This includes how to use multiple tools of teacher evaluation, and both formative and summative evaluation.
- Guidelines for teachers and evaluators of the alternative system should be provided, and there should be regulations and sanctions for evaluators when they ignore any part of the system policy.

Chapter Eight: Discussion and Implications

8.1 Introduction

This study has analysed and evaluated the current teacher evaluation system in Kuwait from the perspectives of teachers, head teachers, and inspectors. The objectives of the research, in this sense, have been to look at the purposes of evaluation, evaluation tools, the involvement of internal and external evaluators, and to consider teachers' views about the extent to which the current system supports them in developing professionally. Furthermore, the study proposed an alternative system for teacher evaluation based on a 'Risk-based analysis' approach and explored the participants' view on its potential for the improvement and development of teacher evaluation in Kuwait.

The study was based on a questionnaire that was distributed to 599 teachers, interviews with nine head teachers, and twelve inspectors, as well as nine focus groups that were conducted with teachers in nine primary schools. The results constitute the basis for a discussion in response to the research questions.

The chapter is divided into sections, as follows: Section 8.2 will discuss the actual and desired purposes of teacher evaluation; section 8.3 will discuss the tools of teacher evaluation that are currently used and that should be used; section 8.4 will discuss the involvement of internal and external evaluators; section 8.5 will discuss the teachers' views about the extent to which current system supports them; section 8.6 will discuss participants' views on the proposed alternative system; section 8.7 will present the implications of the findings and make recommendations; section 8.8 discusses the limitations of the study; and section 8.9 proposes ideas for further research.

8.2 The purposes of teacher evaluation

The current teacher evaluation system in Kuwait, according to the teachers' views, has a stronger focus on determining teacher performance and making decisions about rewards and sanctions than on promoting professional development. Head teachers and inspectors concur with this view. When asked about preferred purposes, while teachers accept that it should be used to make decisions about rewards and sanctions, they favour also using evaluation for professional development purpose. The results show that there are

statistically significant differences between the actual and desired purposes, based on their perspectives ($p < .05$). Head teachers and inspectors support the idea that teacher evaluation should be used for professional development and indeed rate it as of equal importance to its summative purposes. This reflects the teacher evaluation policy, which states that the summative and professional development purposes should have equal weight (KTS, 2010; MOE, 2011). Thus, in both teachers' and evaluators' views, there is an imbalance between the actual and intended purposes of teacher evaluation in Kuwait when it comes to professional development and summative purposes: in practice, there is too little focus on teacher development. This result was unexpected and surprising due to this researcher's expectation of whether the policy of teacher evaluation determines the purpose, evaluators or MOE will be committed to the policy.

This study found, however, some disparity in views within the teacher group. Teachers with long working experience (more than 20 years) tend to claim that practice is more aligned with policy. In other words, that teacher evaluation in practice gives priority to both professional development and summative purposes. There are also some contradictions between the findings of the current study and previous research by Alsanafi (2012). She found that professional development and determining teacher performance are largely met by the current system, but that the system is not appropriate for transferring the latter information into decisions about sanctions and rewards. This may be due to the smaller sample size: Alsanafi's study involved only 110 social science teachers in two educational districts, while the current result came from large sample in three districts, from teachers of several subjects, and included head teachers and inspectors.

There are several possible explanations why participants see the evaluation system as not sufficiently focussing on professional development. First, there is a lack of openness regarding the final individual teacher evaluation report, which is not provided to teachers but kept confidential. Teachers thus do not know about their progress or weaknesses at the end of the school year, and cannot compare their performance overall to the criteria of effective teaching. Second, the system does not recognise any professional development teachers do undertake; except for teachers whose performance is a cause for concern, teachers are not obliged to attend courses designed by the inspection departments or other schools and no records are kept of their attendance at training courses and workshops. Thus, teachers whose performance has been rated as good or outstanding, may not be

motivated to participate in professional development because training course attendance is not accounted for in their evaluation. Third, some teachers are evaluated more on their non-duty teaching activities, which include cooperating with school management and organising school activities, than on the other aspects of their performance. This may lead some teachers to focus on organising school activities or doing some work that teachers do not have to do (e.g. school management tasks) in order to attain an outstanding performance result, rather than trying to improve their teaching or enhance their strengths.

Suggestions with regards to achieving a balance between the two purposes can be inferred from the findings. Firstly, openness with regards to the final evaluation, which in practice would involve giving individuals their final detailed report. In other words, *disclosure* can be used to support professional development. By providing them with the details of their final performance report, teachers would be more aware of what they have achieved and what they need to improve upon (while encouraging outstanding teachers to maintain that level). This also may help them to understand what constitutes effective teaching, and thus know how to deliver the best performance that they can. Providing detailed information in the final report, of course, may cause problems such as less collaboration if a teacher gets a lower score than he or she expects, although many head teachers in this research did not think this would be the case. Contrasting views are reported in Alhamdan (1998) for example, who found that some teachers, head teachers, and inspectors supported keeping results confidential because of the belief that they would cause problems among teachers and between teachers and evaluators (head teachers and inspectors). Similarly, Alsanafi (2012) suggested that the final result of the evaluation should be provided to teachers but without a grade attached (only provide comments about a teacher's performance) to avoid causing problems between teachers and head teachers.

The second suggestion to arise from the findings is that teacher evaluation reports should be used in the designing of training courses and workshops, in order to meet the professional development needs of the teachers, whether addressing areas of weakness or building on their strengths or updating teachers on changes or new pedagogy. As highlighted by Darling-Hammond (2013), teacher evaluation should be linked to professional development opportunities for teachers, as evaluation alone cannot lead to the necessary improvements and development. Albustami (2014) also argued that professional development courses should be based on evaluation reports. Moreover, a record of training

courses attended should be included in teacher evaluation to encourage teachers, especially those who have been evaluated as very good or outstanding, to attend training courses and workshops. Nolan and Hoover (2008) also underline the importance of documenting professional development activities in an effective teacher evaluation system. In addition, schools should receive financial support or have the freedom to collaborate with private agencies to organise their own training courses and workshops, rather than these being organised at central level. This will enable schools to be more responsive to their own cohort of teachers.

Third, there should be explicit and strict sanctions for all teachers, leading to dismissal, when performance is consistently very poor. In order to motivate teachers to develop their practice, in addition to money or promotions, rewards should be available which lead indirectly to professional development, such as scholarships and travel bursaries to attend conferences. Moreover, all teachers should have the same opportunities for promotion; priority should not be given to teachers based on nationality and teacher evaluation outcomes, as this creates a barrier for foreign teachers and means they may not be motivated to perform their best and indeed, may only perform sufficiently well to avoid sanctions.

8.3 The tools of teacher evaluation

Teachers indicated that observation is the most frequently used evaluation tool, while student evaluation is the least used, with three-quarters of the teachers stating that student evaluation is never used. Other tools such as student achievement, teacher portfolios, self-evaluation, and peer evaluation for formative purposes, fall somewhere in between the two extremes. Teachers agreed that observation is the most valuable tool, and there was no significant difference between their actual and preferred choice ($p > .05$). However, they indicated a preference in using other tools more frequently, as a paired t-test result had $p < .05$. Head teachers also stated that observation is the most frequently used tool, and also indicated that teacher portfolios are commonly used, along with self-evaluation and student achievement. Inspectors mentioned observation as a tool that is “always” used and a few of them also mentioned student achievement results and portfolios. There was little discrepancy between what head teachers and inspectors said *is* used and what *should be* used; the most typical pattern was support for using a range of different tools.

Thus, based on this study, the current teacher evaluation system uses classroom observation more frequently than any other tool. This is in line with the teacher evaluation policy, which stipulates using a standardised checklist for determining a teacher's performance inside the classroom (KTS, 2010). When it comes to using other tools, responses from evaluators were quite varied. Surprisingly, what emerges from the findings is that using tools other than observations is a matter of personal initiative in the current system. Furthermore, an analysis of teachers' view is that use of the range of evaluation tools varies according to demographics, educational districts, subjects and teaching experiences, because the evaluators' personal initiatives are different between schools, and from teacher to teacher. These findings contradict an earlier study of the Kuwaiti evaluation system by Sabti (2010) which found that it depended solely on observation to collect evidence about a teacher's performance by head teachers, inspectors, and heads of departments.

There was general agreement that a broad range of tools should be used to evaluate teacher evaluation, such as observation, students' achievement, self and peer evaluation, student evaluation and teacher portfolio. Previous studies have demonstrated that using multiple tools to collect data to evaluate teacher performance is one way to ensure an effective system or improve its effectiveness (Colby et al., 2002; Darling-Hammond et al., 2012a; Stronge, 2006b; Kane & Staiger, 2012). Furthermore, this study's findings are also consistent with Alsanafi's (2012) with regards to the need for teachers to have a greater opportunity to be involved in their evaluation through 'self-evaluation'.

The use of a range of tools can assist evaluators in collecting reliable data about a teacher's performance that will help them to make more valid and fair judgements than they would by simply depending on one tool such as observation. Moreover, by using a range of tools, a teacher may receive better information about their performance than they would from observation alone, and better information will facilitate improvement. These results are in line with studies that consider the benefits of combining multiple tools (Burnett et al., 2012; Lachlan-Haché, 2011; Kane and Staiger, 2012; DePascale, 2012; Hanover Research, 2012).

While the use of multiple tools was favoured in principle, there were mixed views with regards to using student evaluation. For example, some head teachers think students may be too emotional to adequately determine a teacher's performance. Some teachers also argued against using student evaluation, particularly computer teachers, who were less supportive of the idea of using student evaluations than other subject teachers. This result may be explained by the differences in lessons types. Computer teachers use the teaching lab where students carry out tasks individually, while other subjects teachers rely on sharing their experiences and other classroom activities in which the students are more dependent on the teacher 'teacher-centred'. Thus, computer teachers may feel that student evaluation is not as helpful as other tools in determining their performance. Some head teachers, on the other hand, thought that students' views could provide valuable information about teachers that can be used when evaluating their performance, for instance, by asking indirect questions leading to responses that will indicate whether the teachers respect their students and show regard for their learning.

To ensure that multiple tools are used in teacher evaluation, tools should be specifically mentioned in the teacher evaluation policy; this would prevent or reduce the use of personal initiative, which may be seen by teachers as infringing on their rights and therefore may provoke resistance to the use of certain tools in their evaluation that are not explicitly mentioned in the policy, especially given that teacher evaluation is used to judge teachers and make decisions resulting in sanctions or rewards. Making the use of multiple tools mandatory in the policy would also protect evaluators from having their personal initiative used against them, such as by head inspectors who consider it a breach of the regulations, thus affecting their performance evaluation. Moreover, including the use of multiple tools in the policy may make it more likely that they will be used across the schools.

8.4 The involvement of evaluators in teacher evaluation

In the current system, teacher evaluation is conducted by both internal (head of department and head teacher) and external (inspector) evaluators. According to the findings, practice reflects policy, in that every teacher is evaluated by a head of department, head teacher, and inspector (KTS, 2010; MOE, 2011).

In the current system, both head teacher and inspectors hold discussions with teachers after their classroom observations and provide them with written feedback, more than holding discussions before the classroom observation. Heads of departments were stated as being the most likely to have discussions both before and after observation with teachers and they were also more likely to provide written feedback than the other two evaluators (head teachers and inspectors). Teachers rated discussions and written feedback with/from heads of departments as more valuable than with the other two groups of evaluators. There are a number of factors likely to contribute to this.

Firstly, head teachers have a huge number of teachers to observe and therefore will struggle to make time for several classroom visits in the school year. In addition, their other responsibilities, such as managing the school, affect the extent to which they fulfil their role as evaluators and thus the value of their role. Indeed, this study also found that the role of head teachers and its value differ from school to school from the teachers' view, because head teachers' workload and responsibilities are probably different. Heads of departments, by contrast, have only a few teachers to observe and most of their responsibilities are within the department.

Inspectors have an even larger numbers of teachers to observe (ranging between 70 and 260 teachers for every inspector) and have other significant responsibilities, such as preparing tests, content and curriculum, meeting with department heads, supervising activities in schools, and designing training. Due to workload, inspectors rarely observe a full lesson and are not able to observe teachers with any regularity; this affects the extent to which they are able to fulfil their role as evaluators and consequently their value to teachers. This study supports research by Weisberg et al. (2009) and Albustami (2014) that for a number of reasons, teacher evaluation often fails to provide accurate and credible information about a teacher's performance. One of these reasons is that when evaluators conduct infrequent and brief observations, they may be inattentive to teachers' performance.

Second, a lack of subject knowledge may affect the extent to which teachers value feedback from head teachers. For example, some head teachers were Arabic teachers and yet they were evaluating maths or English teachers, despite knowing no English and not having a good knowledge of maths. Albustami (2014) also found that an evaluator's

subject knowledge plays a significant role in teacher evaluation and that lack of subject knowledge will undermine and even invalidate the judgements they make about a teacher's performance. This is also confirmed by this study, since there was a statistical difference between teachers' views, based on subject, regarding the head teachers' role and their value as evaluators. English teachers saw head teachers as less involved and rated their role as less valuable than other subject teachers that taught in Arabic. However, head teachers were able to evaluate the teachers to some extent in those areas of transferrable knowledge such as pedagogy and classroom management, and other important elements which are common across subjects, such as managing behaviour, motivating students and building learning environments, as pointed out by Hill and Grossman (2013). Furthermore, head teachers were able to compensate for their lack of subject knowledge by asking the heads of departments to attend the observation and highlight any issues that they perceived.

Third, the lack of training for head teachers and inspectors prevents them from keeping up-to-date with innovations in the field of education and affects the value of their role. For example, if an evaluator left teaching more than 10 years ago then they would need to be retrained in current teaching practices and evaluation in order to ensure the usefulness and accuracy of their evaluations. Albustami's (2014) research also found that one of the reasons that undermined the value of teacher evaluation was that the evaluators were sometimes not trained well enough to conduct teacher evaluation. Therefore, several researchers have argued that training courses should be targeted to ensure that evaluators are properly trained (Darling-Hammond et al., 2012b; Albustami, 2014; Partee, 2012).

Fourth, a lack of experience or specialisation at school level could also be seen as affecting the ability of external evaluators (inspectors) to evaluate teachers, and thus may make their role seem less valuable than other evaluators' role. For example, an inspector with experience of working in a high school may have difficulties evaluating teachers in a primary school, since they may have little idea about how teachers should teach and deal with students at this level. This study, however, found that the head of department's experience in evaluating teacher also could negatively affect the extent to which their evaluation was perceived as valuable. The study, for example, found a statistical difference between the teachers of computing and teachers of other subjects regarding the value of the department heads. This is probably due to the fact that some schools do not have a head for the computer department so they ask an experienced teacher to be a substitute head of

department (As written by some of computer teachers in the questionnaire "we do not have a head of department"). This person may be an experienced teacher but he or she is likely to have little experience of evaluating teachers.

Fifth, gender was found to an important factor affecting the role and perceived value of the inspector. An example is in the area of Islamic studies. When evaluating the teaching of the Quran, evaluators look at how words are pronounced, a part of teaching the Quran in Arabic, which is called 'Tajweed'. Where female teachers' faces are covered by a veil, male inspectors encounter difficulties evaluating 'Tajweed'; the same issue may be encountered in English and Arabic, as inspectors have to look at how words are pronounced. Male inspectors thus cannot accurately evaluate performance and therefore the feedback they provide to female teachers may be of less value than that provided by the female head of department.

8.5 Extent to which the current system supports teachers

The study found that in the current system, determining performance and making decisions about rewards and sanctions are the main purposes of teacher evaluation. The study also found that teacher evaluation mostly depends on observation, and that teachers obtain feedback and hold discussions with evaluators. The question remaining is the extent to which the current system supports teachers in developing their performance, in their view.

Results varied, with some teachers indicating that the current teacher evaluation system did not help them in developing their performance, while other teachers stated that it did. In the current system, there is both support and lack of support in the various aspects of teaching: understanding of the content being taught; use of pedagogies; lesson planning; understanding of what constitutes effective teaching; identifying weaknesses and determining strengths; organisation of activities in the classroom; improving the teacher's ability to deal with student discipline and behaviour problems; improving the teacher's ability to motivate students in terms of their learning; improving a teacher's ability to deal with individual differences between students; affecting teachers' continuous assessment of student learning; and affecting teachers' ability to provide students with effective feedback.

The findings must be approached with caution because they cannot be extrapolated to apply to all teachers. Teachers' views seemed to be split based on length of experience: teachers with more experience viewed the current system as supporting them in developing their performance when compared to teachers with less experience who did not believe that the system supports them in this regard. The greatest mean difference was found between teachers with less than 10 years' experience and teachers with more than 20 years' experience. This result was surprising as the expected result was that teachers with less experience might see the current system as more supportive of them than other teachers, as they do not have enough experience in teaching; therefore, they would see teacher evaluation as supporting them in following the right way and making their teaching more effective.

One possible explanation for this is that a teacher with more than 20 years of experience in teaching may base their views on teacher evaluation on their past experience of working under two different systems (the previous system and the current system). In other words, teachers with less than 10 years of experience may only reflect on the current teacher evaluation system (no. 36/2006), while teachers with more than 20 years' experience may be comparing the two systems (no. 461/1993 and 36/2006), implying that teachers with many years experience may perceive the current system as better than the previous one with regard to the development of performance.

It may also be that teachers with more experience are better able to interpret and derive benefits from the feedback with which they are provided in order to develop their performance than those with less experience, who may find the insufficient disclosure of information in the current system an obstacle to developing their teaching. In other words, more experience may help teachers to understand the feedback and thus use it to develop their performance. Indeed, this study found that the value of the inspectors' feedback differs between groups of teachers, as teachers with more experience rated the inspectors' role more valuable than those with less experience. On the other hand, Tuytens and Devos (2012) found that teachers with more experience saw feedback as less useful than teachers who were less experienced. Delvaux et al. (2013) also found that among teachers with limited years of experience, feedback was positively related to the perceived effects of teacher evaluation on professional development.

What the current system is perceived as doing well is recognising teachers for their hard work, since teachers are generally rewarded for an outstanding performance. The results from teachers show that teacher evaluation leads to monetary rewards, i.e. in the form of an annual bonus or increase in salary. The results also show that teacher evaluation is supportive of teachers in terms of promotions, i.e. becoming head of department. However, there is a difference between female teachers (m=3.50) and male teachers (m=3.09) in this regard. The difference is probably caused by the fact that male teachers have less opportunity for promotion since there are fewer primary school posts for male teachers in Kuwait, because it is the MOE's policy that primary education should be delivered by female teachers. This suggests that the current teacher evaluation system may motivate some teachers to develop their performance but not others. Delvaux et al.'s findings (2013) concur to some extent with the findings of this research in that their study showed that a teacher evaluation system whose purpose is summative has little, but positively significant, effect on teachers' development. They suggested the reason might be that teachers feel under pressure when the purpose of evaluation is summative and may feel compelled to undertake professional development. However, this should be seen in the context of the system investigated in their study, which was used for professional development rather than for summative purposes.

8.6 An alternative system based on a risk-based analysis approach

Regarding the implementation of an alternative system, the vast majority of participants (both evaluators and teachers) were supportive, a surprising and unexpected result indicating that participants viewed the alternative system as meeting their expectations more effectively than the current system, perhaps because it suggests providing teachers with mid and final year reports, the use of multiple tools, and consider the RBA approach.

The alternative system would firstly, improve *validity and reliability* in teacher evaluation. First, it would implement multiple tools in evaluating teachers' performance, leading to more accurate evaluations of their performance as evaluators will get a more comprehensive picture that includes activities both inside and outside the classroom, and takes into account non measurable factors by allowing teachers, students, and teachers' work (portfolios) to form part of evaluation. Furthermore, the use of a standardised checklist for self-evaluation as it would also ensure that all aspects of teacher performance are addressed, preventing teachers from evaluating some aspects and not others. The

standardised checklist would also help evaluators analyse and compare a teacher's performance with the criteria of teacher evaluation.

Second, attaching student standardised test results to the final individual performance evaluation reports is one way of making an appropriate decision about a teacher's performance. This study found that by including student achievement data in teacher evaluations, the degree to which teachers have been able to facilitate students' learning can be compared with their performance reports to make decisions about their performance. However, participants felt it was crucial that standardised tests be created and analysed by a neutral party of experts to ensure accurate results. Furthermore, in analysing the results, certain factors would need to be taken into account, for example when teachers move to another school in the middle of the school year and another teacher takes their place. Similarly, account needs to be taken of the effect on test results of students who have special educational needs and students whose poor language skills impact on their achievements in other subjects.

The third way in which the alternative system would improve the accuracy of evaluation is by attaching other supporting documentation to the final individual teacher evaluation report, consisting of: reports about certificates of attendance for training courses and workshops; official certificates of appreciation from inspection departments or educational districts or state institutions; reports on a teacher's attendance and absences; warning letters for late attendance; and any certifications that teachers have obtained in the school year with regard to teaching or learning, even if obtained from private institutions or centres, as part of their personal professional development. Findings from this study suggest that these documents will give a more comprehensive picture of a teacher's efforts regarding professional development and will also provide evidence of a teacher's commitment to teaching, in order to help inspectors make fair decision about a teacher's performance.

Fourth, in the alternative system, all teachers are evaluated by internal evaluators before risk detection is carried out. The internal evaluators would be expected to be able to make accurate judgements about teachers' performance; head teachers and heads of departments are with teachers on a daily basis so are in a better position to conduct evaluations and

follow-up discussions, as opposed to the inspectors who only visit three or four times in a school year and observe the teachers for brief periods of time.

Fifth, by removing their direct involvement in teacher evaluation, external evaluators would be in a position to focus on risk detection. Their role would be to look at the final teacher evaluation reports, student outcomes and other documents, to evaluate a teacher's professionalism, looking for any discrepancies between multiple sources of information and any effects of personal relationships on the evaluation that might reduce its validity. Also, risk detection would be conducted by a committee of inspectors, thus further reducing subjectivity and increasing the validity of the decisions, as they would be based on multiple perspectives.

In the alternative system, teacher performance would be linked with *promotions and rewards*, but also *sanctions* would be introduced for those who underperform. Teachers would be rewarded and promoted when evaluated as outstanding, thus gaining recognition for their hard work. On the other hand, sanctions would be applied after a teacher has been given the opportunity to improve, through tailored and intensive evaluation, and timely decisions would be made where necessary to remove teachers who are not able to teach students.

The alternative system would also *facilitate* and *promote development and improvement.* Teachers would be given a mid-year report and feedback after observation, which would encourage and motivate them to work towards improving themselves early on in the school year (second semester), as it was under the previous system in Kuwait. As highlighted by Alhamdan (1998), teachers found the mid-year report helpful in improving their performance in the second half of the term. Findings from this study also suggest that a mid-year report would encourage teachers to attend courses, whether provided by private institutions, teachers' unions or the inspection department, to improve or develop their teaching. Moreover, a mid-year report would also help internal evaluators to identify weak performances early on and plan suitable interventions, e.g. by nominating them to attend training courses.

Teachers would also be given a detailed final evaluation report, which would give them the opportunity to know exactly what they need to improve or develop, since they would know what aspects of their performance are being evaluated. This would help them aspire to becoming better teachers, thus improving their students' learning. Participants in the study indicated that they would wish the report to be confidential, since it could contain sensitive information, and to protect evaluators from being pressured by teachers to change the report. This could be addressed by sending the reports directly to the teacher's home or making them available for collection from the educational district centre after being signed off and accredited by the evaluators at the end of the school year. However, this method of delivery would not necessarily prevent discussions between teachers about performance, since discussing results with colleagues and helping each other does not relate to whether teachers know their colleagues' reports.

Furthermore, this alternative system would document teachers' professional development, which may increase motivation to attend training courses and workshops, especially the motivation of very good teachers and outstanding teachers, as a record of training courses undertaken would be attached to the evaluation report. Thus a teacher's efforts to develop professionally would be acknowledged and distinguished from those teachers who make little effort. Likewise, inspectors could prepare training courses for all teachers based on the final individual evaluation reports and other evidence that is provided to them. Inspectors will be in a better position to determine which training courses are most needed.

A tailored evaluation would also improve teacher performance: since inspectors would not be evaluating all teachers, they would be able to focus on those teachers considered at risk, would be provided with all relevant information before evaluating them and therefore would not waste time determining the teacher's strengths and weaknesses. They would then be able to follow through with appropriate interventions designed to address the risks identified, beginning where the internal evaluators stopped. If the teacher at risk does not improve, they would also receive intensive evaluation by two inspectors to improve his/her performance. The expected benefit is that there will be more effective improvements in the teacher's performance.

Moreover, since inspectors will be involved in directly evaluating teachers at risk, the number of teachers will be significantly less and therefore the head inspector will be able to match an inspector to a teacher in two significant ways: specialisation in the same school level and gender. These are two areas identified as problematic and as reducing the value and effectiveness of the inspector's role as evaluator. The head inspector may ensure that female inspectors are with female teachers and male inspectors are with male teachers, and that those inspectors also have the appropriate subject specialisation and experience of school level, and will therefore be able to offer a tailored and intensive evaluation to improve teachers' performance.

To sum up, the alternative system would serve both professional development and summative purposes, as desired by participants in the present study, and argued by numerous studies as leading to a more effective system (Colby et al. 2002; Stronge 2006b; Peterson and Comeaux 1990). In the alternative system, teacher performance is evaluated using multiple tools, since both this study and others (Colby et al., 2002; Darling-Hammond et al, 2012a; Stronge, 2006b; Kane & Staiger, 2012) have found a range of tools to be necessary. In the alternative system, both external and internal evaluators are involved in teacher evaluation as internal and external evaluation is necessary but there is a better division of labour between them. As pointed by Nevo (2001, p. 101), each can learn something from the other, but any "evaluation (internal and external) has to be modest, acknowledging its limitations". In the alternative system, teachers at risk will receive tailored and intensive evaluation. As Nolan and Hoover (2008) suggested, in any well-designed system of teacher evaluation, the procedures should differ between high-performing teachers and low-performing teachers in order to give them better direction, guidance and support to improve. Therefore, the alternative system has the potential of making better use of teacher evaluation in Kuwait's education system.

8.7 Implications

A number of implications for researchers and decision makers arise from the findings of the research, of particular pertinence to the Kuwaiti context.

This research clearly shows stakeholders' perspectives regarding the benefits and conflicts of the current teacher evaluation system. According to the MOE (2013c), it has set out a plan to develop education in Kuwait. One component of the development of education in Kuwait is evaluation and measurement. Therefore, the results of this study might help decision makers in the Kuwaiti MOE by providing in-depth information about the current teacher evaluation system from the perspective of its key players, articulating their views and concerns and using them to propose changes that would develop the teacher evaluation system in the Kuwaiti context.

The main recommendation is that Kuwait's MOE should consider preparing and conducting the alternative system with consideration of risk-based analysis as a principle for teacher evaluation to overcome problems and address the needs identified by teachers and evaluators, as proposed in Table 8.1. In conducting the alternative system, the MOE should take into account the following:

A. Evaluators should have appropriate training and access to workshops about how to conduct formative/summative evaluation and the tools of teacher evaluation. Training and workshops also should be provided for evaluators on how to implement the alternative system.
B. Training and workshops should be provided to teachers to prepare them to conduct self-evaluation and peer evaluation, as well as how to prepare a teacher portfolio.
C. The number of inspectors needs to be increased so that they can focus on teachers alongside their other responsibilities.
D. There should be official guidelines (a booklet) for teachers and evaluators of the alternative system, to use as a reference.
E. There should be regulations and sanctions for evaluators if they ignore any parts that are included in the policy of the alternative system.

Table 8.1: The Proposed alternative system for teacher evaluation (Amended)

Steps	Producers
First step: Individual teacher performance evaluation	1) Each teacher is evaluated by the head teacher and head of department every school year (from September to May). 2) Each teacher is provided with a mid-year report and feedback after observation. 3) Each teacher is provided with the final report in detail at the end of the school year after being signed off and accredited by the evaluators. The report is sent to the teacher's home after school, or collected from the educational district offices. 4) The final reports of the individual evaluation are linked with the teacher's salary scale, annual bonus, promotions, and sanctions, such as dismissal or redeployment to a non-duty teaching job. 5) Evaluators have to use observation, student evaluations (survey or interview), self-evaluation (standardised checklist), and peer evaluation for formative purpose, along with a teacher portfolio. 6) Standardised test results are attached to the individual teacher evaluation report. These tests should be designed and analysed by the National Centre of Educational Development in the MOE. School can be involved in administering these tests. These tests come with some caveats that should be considered, as they may have negative effects on both students and teachers. 7) Other documentation 'signals' are attached: certificates of attendance at training courses and workshops; reports about a teacher's attendance and absences, warning letters for late attendance, and any certifications that teachers have obtained in the school year with regard to teaching or learning, even if teachers have obtained them from private institutions or centres. Note: • Head teachers can ask for help from assistant head teachers and inspectors can participate in formative assessments if needed. • For students' evaluation: evaluators should select some, not all classes, taught by a teacher (the evaluator has the freedom to choose the number). For example, teachers who teach more than five or six classes can have student evaluations from three classes. • A standardised test might be expensive, so the MOE may use students' results in achievement tests.

Table 8.1 (Cont.)

Steps	Producers
Second step: Risk detection	This step should start at the end of the school year, in May or June. Inspection departments for different subjects in six educational districts will receive individual teachers' evaluation reports, standardised tests, and other documentation. The inspection departments for each subject in each educational district will arrange for a committee of inspectors to initiate risk detection. Inspectors determine whether or not there is risk, based on the evidence above. Risk can be detected when there is a difference between the teacher evaluation results, standardised tests results, and other documentation (even if they are outstanding or very good), or if a teacher is evaluated as weak or good. If the inspectors determine that there is no risk in the teacher's performance, the teacher will receive regular evaluation (if no risk is also identified in the next annual risk detection). If the inspectors detect a risk in the teacher's performance, they go through the history of that individual teacher's evaluations from the previous years. If a teacher was evaluated as good or weak in one year out of a certain number (e.g. five or ten), inspectors will start to tailor this teacher's evaluation in the next academic year (starting in September). If the teacher is evaluated as very good or outstanding for several years, the teacher will receive regular evaluation (if no risk is also identified in the next annual risk detection).
Third step: Tailored, intensive and regular evaluation	1) Tailored evaluation: the teacher is individually evaluated by the head teacher and one inspector for one school year. 2) Intensive evaluation: teacher is evaluated for another school year by the head teacher and two inspectors, if the teacher does not improve in tailored evaluation. 3) Regular evaluation: all teachers (except teachers in intensive evaluation) are evaluated by an inspector, head of department, and head teacher every three school years. Regular evaluation will be the same as in the first step: individual teacher evaluation but with inspectors' participation. Note: - In tailored and intensive evaluation, inspectors will focus on evaluating teaching and learning aspects, while the head teacher will focus on the teacher's commitment to their work and their collaboration with colleagues and school staff. - Two subject inspectors should be free to evaluate teachers in intensive evaluation in each district, while all inspectors participate in regular evaluation. - Tailored and intensive evaluation will be the same as in the first step, i.e. individual teacher evaluation as has been described, but carried out by the inspectors instead of the department head.

Furthermore, the key recommendations, based on the literature and empirical research conducted by this researcher, are summarised as follows

A. Teacher evaluation should be used for professional development and summative purposes.
B. Training courses should be linked with teacher evaluation reports. These courses should be organised by MOEs/organisations and schools should also have the freedom to design their own training courses.
C. Teachers should be evaluated using various tools (observation, student achievement, self-evaluation, peer evaluation for formative reasons, student evaluation, and teacher portfolios) so that evaluators can acquire more accurate data about a teacher's performance.
D. Teacher evaluation should include both internal and external evaluators, as each can learn something from the other.
E. There should be openness with regards to teacher evaluation or disclosure, which in practice would involve giving individuals their detailed report.
F. The teacher evaluation report should include documents or a record of the teacher's attendance at training courses, workshops, conferences, etc. to motivate the teacher to be more involved in professional development.
G. A teacher's evaluation history should be considered in the teacher evaluation system since the aim of the procedure is to differentiate between teachers that are outstanding and those that are not.

8.8 Limitations of study

The strengths of this research are that it involves a large sample of teachers, head teachers, and inspectors in one study, and findings are based on both qualitative and quantitative data through the use of questionnaires, interviews, and focus groups. However, it also has a number of limitations: firstly, the study was confined to specific participants in primary schools, as the nature and purpose of this study is focused on this group. Time constraint was the reason for not expanding the research to include other schools, namely middle and high schools. Second, this study was geographically restricted to the educational districts of Ahmadi, Asimah and Farwaniya in Kuwait, which represent three out of six educational districts in the country. This researcher believes that these districts are representative of conditions in Kuwait since because of the MOE's centralised operation, the teacher evaluation system is uniformly implemented across the schools. Third, this researcher has

excluded parents from this study because teachers, head teachers and inspectors are in a better position to provide data regarding the current teacher evaluation system and assess whether the idea being developed is likely to work and be valid, whereas parents have no involvement in the current system. Furthermore, involving parents in this research would have exposed the study to higher costs, required more time and might possibly have encountered a lack of collaboration.

8.9 Suggestions for further research

The current teacher evaluation system in Kuwait is under researched. It is therefore suggested that further research be undertaken to examine the following:

A. Review the teacher evaluation criteria from the perspectives of teachers and evaluators in different educational districts in order to determine which criteria are still appropriate and which are in need of further development or modification.
B. Determine which combination of evaluation tools is appropriate from the perspectives of teachers, head teachers and inspectors (such as holistic, numerical, and portfolio or matrix approaches).
C. Assess the quality of the training courses and workshops that are provided either by schools or through inspection departments, from the teachers' perspectives.
D. Finally, if the MOE applies the proposed alternative system, which is based on an RBA approach, in educational districts or if the MOE implements it in some districts as a pilot, research will need to be conducted to determine how effective the alternative system is for teacher evaluation in the Kuwaiti context in practice.

References

Airasian, P., & Gullickson, A. (1997). *Teacher self-evaluation tool kit*. Thousand Oaks, CA: Corwin press.

Alaani, N., Maqdad, M., & Aldousari, R. (2003). *Evaluation, measurement and tests*. Kuwait: Arab Open University.

Albustami, G. (2014). Improving the teacher's evaluation methods and tools in Abu Dhabi schools- Case study. *Athens Journal of Social Sciences*, *1*(4), 261-273.

Alhamdan, J. (1998). Teacher evaluation in the State of Kuwait. *Educational Journal: Kuwait University, 12*(47), 289-312.

Alkhayat, A., & Dhiab, A. (1996). Teacher evaluation system in the Ministry of Education in Kuwait: Evaluation study. *Educational Journal: Kuwait University, 10*(38), 27-78.

Almutairi, T., Tymms, P., & Kind, P. (2015). The tools of teacher evaluation: What should be used in teacher evaluation from the teachers' perspective. *2015 International business and education conferences proceedings, London*, *June*, 326.

Alnajar, N. (2010). *Evaluation and measurement with applying SPSS*. Jordan: Dar-Alhamed.

Alsanafi, S. (2012). The effectiveness of teacher evaluation system from social studies teachers' perspectives in the middle school in Kuwait. *Educational Journal: Kuwait University*, *45*(1), 309-365.

Altrichter, H & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement?. *School Effectiveness and School Improvement*, *26* (1), 32-56.

Arnodah, I. (2013). Teacher improvement through peer teacher evaluation in Kenyan schools. *European Journal of Training and Development, 37*(7), 635-645.

Askar, A., Jamea, H., Alfarra, F., & Hawana, W. (2009). *Introduction to scientific research: Educational, psychological and social*. Kuwait: AlFalah Library.

Attinello, J., Lare, D., & Waters, F. (2006). The value of teacher portfolios for evaluation and professional growth. *NASSP Bulletin*, *90*(2), 132-152.

Baker, E., Barton, P., Darling-Hammond, l., Haertel, E., Ladd, H., Linn, R., Ravitch, D., Rothstein, R., Shavelson, R., & Shepard, L. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from http://www.epi.org/publication/bp278/.

Best, J., & Kahn, J. (2006). *Research in education* (10th ed.). Boston: Pearson Education.

Biesta, G. (2012). Mixed methods. In J. Arthur, M. Waring, R. Coe & L. Hedges (Eds.), *Research methods and methodologies in education*. London: SAGE.

Black, P., & Wiliam, D. (2006). Assessment for learning in the Classroom. In J. Gardner (Ed.), *Assessment and learning.* London: SAGE Publication Ltd.

British Educational Research Association. (2011). *Ethical guidelines for educational research*. Retrieved from https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1

Brix, J., Grainger, P., & Hill, A. (2014). Investigating mandatory peer review of teaching in schools. *Australian Journal of Teacher Education, 39*(4), 83-99.

Brown, J. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.

Burnett, A., Cushing, E., & Bivona, L. (2012). *Uses of multiple measures for performance-based compensation*. Washington, DC: Center for Educator Compensation Reform. Retrieved fromhttps://www.tifcommunity.org/sites/default/files/35044_CECR_MultipleMeasures_508_updated.pdf

Burns, R. (2000). *Introduction to research methods* (4th ed.). London: SAGE.

Burr, B. (2015). *Student Voices in Teacher Evaluations* (Doctoral dissertations). Retrieved from ProQuest Dissertations & Theses Database (Order No. 3702904).

Burton, D., & Bartlett, S. (2009). *Key issues for education researchers*. London: SAGE.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research, 6(1),* 97-113.

Bryman, A (2012). *Social research methods* (4th ed.). Oxford: Oxford University Press.

Cai, Y., & Lin, C. (2006). Theory and practice on teacher performance evaluation. *Frontiers of Education in China, 1*(1)*,* 29–39.

Check, J., & Schutt, R. (2012). *Research method in education*. London: SAGE.

Chorrojprasert, L. (2005*). The use of teaching portfolios by secondary school teachers in Thailand*. (Doctoral dissertation, University of Wollongong, Australia). Retrieved from http://ro.uow.edu.au/theses/437/

Coakes, S., & Steed, L. (2009). *SPSS analysis without anguish: Version 16.0 for windows*. Milton-Brisbane: Wiley.

Cohen, L., Manion, L., & Morrison, K. (2007) *Research method in education* (6th ed.). London: Routledge.

Colby, S., Bradshaw, L., & Joyner, R. (2002). *Teacher evaluation: A review of the literature*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Cooper, D., & Schindler, P. (2001). *Business research methods* (7th ed.). New York: McGraw Hill.

Creswell, J. (2012) *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.) Boston: Pearson Education.

Cronin, J., Kingsbury, G., McCall, M., & Bowe, B. (2005). *The impact of the No Child Left Behind Act on student achievement and growth*. Portland, OR: Northwest Evaluation Association.

Cronin, L., & Capie, W. (1986). *The influence of daily variation in teacher performance on the reliability and validity of assessment data*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012a). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), p.8-15.

Darling-Hammond, L., Cook, C., Jaquith, A., & Hamilton, M. (2012b). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford, California: Stanford Centre for Opportunity Policy in Education. Retrieved from https://edpolicy.stanford.edu/sites/default/files/publications/creating-comprehensive-system-evaluating-and-supporting-effective-teaching_1.pdf

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* San Francisco: Jossey-Bass.

Davidson, M., Jensen, B., Klieme, E., Vieluf, S., & Baler, D. (2009). *Creating effective teaching and learning environments: First results from TALIS*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/edu/school/43023606.pdf

Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, *36*(2013), 1-12.

deMarrais, K. (2004). Qualitative interview studies: Learning through experience. In K. deMarrais & S. Lapan (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences*. London: Lawrence Erlbaum Associates, Publishers.

Denner, P. R., Miller, T. L., Newsome, J. D., & Birdsong, J. R. (2002). Generalizability and validity of the use of a case analysis assessment to make visible the quality of teacher candidates. *Journal of Personal Evaluation*, *16*(3), 153-174.

DePascale, C. (2012). Managing multiple measures. *Principal, 91*(5) 6-10.

Dinham, S., & Scott, C. (2003). Benefits to teacher of the professional learning portfolio: A case study. *Teacher Development*, *7*(2), 229-244.

Donaldson, M., & Peske, H. (2010). *Supporting effective teaching through teacher evaluation: A study of teacher evaluation in five charter schools*. Washington, DC: Center for American Progress. Retrieved from http://www.americanprogress.org/issues/2010/03/teacher_evaluation.html

Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Manitoba: Manitoba Education, Citizenship & Youth.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *Bmj*, *324*(May), 1-9.

Ehren, M., & Honingh, M. (2011). Risk-based school inspection in the Netherlands: A Critical reflection on intended effects and causal mechanisms. *Studies in Educational Evaluation, 37*(2011), 239-248.

Ehren, M. (n.d.). Risk-based school inspections of Dutch school board: A critical reflection on intended effect and causal mechanisms. *School Inspection Project - European Commission*. Retrieved November 13, 2012 from http://schoolinspections.eu/wp-content/uploads/downloads/2012/05/Netherlands_PT.pdf

Ehren, M., & Swanborn, M. (2012). Strategic data use of schools in accountability systems, School Effectiveness and School Improvement. *An International Journal of Research, Policy & Practice*, *23*(2), 257-280.

Eid, G. (2005). Teacher evaluation in high School in Kuwait: A comparative study of self-evaluation, student evaluation and heads of departments. *Educational Journal: Kuwait University, 19*(76), 79-149.

Elst, H. (2013). *Foundations of descriptive and inferential statistics*. Karlsruhe: Karlshochschule International University.

Eri, R. (2014). Peer observation of teaching: Reflections of an early career academic. *Universal Journal of Educational Research, 2*(9), 625-631.

Faubert, V. (2009). School evaluation: Current practices in OECD countries and a literature review. *OECD Education Working Papers*, No. 42. Retrieved from http://dx.doi.org/10.1787/218816547156

Ferguson, F. (2010). *Student perceptions of teaching effectiveness: Discussion brief*. Cambridge: National Center for Teacher Effectiveness and the Achievement Gap Initiative at Harvard University.

Figlio, D., & Loeb, S. (2011). School accountability. In Hanushek, E., Machin, S., & Woessmann, L (Eds.), *Handbooks in economics*: *Volume 3*. The Netherlands: North-Holland.

Flick, U. (2009). *An introduction to qualitative research* (4th ed.). London: SAGE.

Fraenkel, J., Wallen, N., & Hyun, H. (2012). *How to design and evaluate research in education* (8th ed.). New York: McGraw-Hill.

Gardner, J. (2006). Assessment and Learning: An Introduction. In Gardner, J (Ed.), *Assessment and learning*. London: SAGE Publication Ltd.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, DC: National Comprehensive Centre for Teacher Quality. Retrieved from http://files.eric.ed.gov/fulltext/ED521228.pdf

Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: focusing on improving teaching and learning*. Washington, DC: National Comprehensive Centre for Teacher Quality. Retrieved from http://files.eric.ed.gov/fulltext/ED532775.pdf

Goe, L., & Croft, A. (2009). *Methods of evaluating teacher effectiveness*. Washington, DC: National Comprehensive Centre for Teacher Quality. Retrieved from http://files.eric.ed.gov/fulltext/ED543666.pdf

Goe, L., Holdheide, L., & Miller, T. (2014). *Practical guide to designing comprehensive teacher evaluation systems: A tool to assist in the development of teacher evaluation Systems.* Washington, DC: Center on Great Teachers and Leaders. Retrieved from http://files.eric.ed.gov/fulltext/ED555655.pdf

Gorard, S. (2008). *Quantitative research in education: Volumes 1 to 3*. London: SAGE

Gurr, D. (2007). Diversity and progress in school accountability systems in Australia. *Educational Research for Policy and Practice*, *6*(3), 165-186.

Greene, J. Caracelli, V. and Graham, W. (1989) Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, *11* (3), 255-274.

Hamilton, L., Berends, M., & Stecher, B. (2005). *Teachers' responses to standards-based accountability*. California: RAND Corporation.

Hansen, M., Lemke, M., & Sorensen, N. (2014). *Combining multiple performance measures: Do common approaches undermine districts' personnel evaluation system.* Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research. Retrieved from http://files.eric.ed.gov/fulltext/ED553415.pdf

Hanover Research (2012). *Best practices for including multiple measures in teacher evaluation*. Washington, DC: Hanover Research. Retrieved fromhttp://www.hanoverresearch.com/wp-content/uploads/2012/05/Best-Practices-for-Including-Multiple-Measures-in-Teacher-Evaluations-Membership.pdf

Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning*. London: SAGE Publication Ltd.

Harris, D. (2013). *How might we use multiple measure for teacher accountability*. Stanford, California: Carnegie Foundation for the Advancement of Teaching. Retrieved from http://files.eric.ed.gov/fulltext/ED560146.pdf

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 372-384.

Hopkins, K., Hopkins, B., & Glass, G. (1996). *Basic Statistics for the Behavioral Sciences* (3rd ed.). London: Allyn and Bacon.

Hounsell, D. (1997). Understanding teaching and teaching for understanding. In Marton, F., Hounsell, D., & Entwistle, N. (Eds.), *The Experience of learning*: *Implications for teaching and studying in higher education.* Edinburgh: Scottish Academic Press.

Isore, M. (2009). Teacher evaluation: Current practices in OECD countries and a literature review. *OECD Education Working Papers*, No. 23. Retrieved from http://dx.doi.org/10.1787/223283631428.

Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches* (3rd ed.). London: SAGE.

Johnson, R., Onwuegbuzie, A., & Turner, L. (2007). Toward a definition of mixed method research. *Journal of Mixed Methods research*, *1*(2), 112-133.

Johnson, S., & Fiarman, S. (2012). The potential of peer review. *Educational Leadership*, *70*(3), 20-25.

Jones, M., Jones, B., and Hargrove, Y. (2003). *The Unintended consequences of high-stakes testing*. Lanham, MD: Rowan and Littlefield Publishers.

Joshua, M., Joshua, A., Bassey, B., & Akubuiro, I. (2006). Attitude of Nigerian secondary school teachers to peer evaluation of teachers. *Teacher Development: An international journal of teachers' professional development, 10*(3), 331-341.

Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. MET Project, Bill & Melinda Gates Foundation. Retrieved from http://files.eric.ed.gov/fulltext/ED540960.pdf

Kellaghan, T., Stufflebeam, D., & Wingate, L. (2003). Introduction. In T. Kellaghan & D. Stufflebeam (Eds.), *International handbook of education evaluation*, 1-6. Dordrecht: Kluwer Academic Publishers.

Kelly, S. (2014). *A case study examining teacher responses to principal feedback of class observations.* (Doctoral dissertation, Kennesaw State University, Kennesaw, USA). Retrieved from http://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1000&context=eduleaddoc_etd

Khachatryan, E. (2015). Feedback on teaching from observations of teaching: what do administrators say and what do teachers think about it?. *NASSP Bulletin*, *99*(2), 164-188.

Knupfer, N., & McLellan, H. (1996). Descriptive research methodologies. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology.* New York: Macmillan

Kuwait Central Statistical Bureau. (2013/2014). *Annual statistical abstract*. Kuwait: Kuwait Government.

Kuwait Constitution. (2008). *The Constitution of the State of Kuwait*. Kuwait: Ministry of Information.

Kuwait Teacher Society. (2010). Guidance and the policy of teacher evaluation: Number 36/2006. *Kuwait Teacher Society Journal*, *June*(1586).

Kuwait Teacher Society. (2012). *Cadre of Teachers No. 28/2011*. Kuwait: Kuwait Teacher Society.

Lachlan-Haché, L. (2011). *Using multiple measures in comprehensive teacher evaluation system*. Washington, DC: American Institutes for Research. Retrieved from http://www.nga.org/files/live/sites/NGA/files/pdf/1107TEACHERLACHLANHACHE.PDF

Ladd, H. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review*, *18*(1), 1–16.

Leo, S., & Lachlan-Haché, L. (2012). *Creating summative educator effectiveness score: Approaches to combining measures*. Washington, DC: American Institutes For Research. Retrieved from http://educatortalent.org/inc/docs/Creating%20Summative%20EE%20Scores_FINAL.PDF

Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. Washington, DC: National Comprehensive Centre for Teacher Quality. Retrieved from http://files.eric.ed.gov/fulltext/ED543776.pdf

Liu, S., & Teddlie, C. (2005). A follow-up study on teacher evaluation in China: Historical analysis and latest trends. *Journal of Personnel Evaluation in Education*, *18*(4), 253-272.

Looft, K. (2002). *Teachers' perceptions of the evaluation process* (Doctoral dissertation, University of Georgia, Athens, USA). Retrieved from https://getd.libs.uga.edu/pdfs/looft_karen_g_200208_edd.pdf

MacBeath, J., & McGlyn, A. (2002). *Self-evaluation: What's in it for schools?*. London: Routledge Falmer.

Mack, N., Woodsong, C., MacQueen, K., Guest, G., and Namey, E. (2005). *Qualitative research Methods: A Data Collector's Field Guide*. North Carolina : U.S. Agency International Development. Retrieved from https://www.fhi360.org/sites/default/files/media/documents/Qualitative%20Research%20Methods%20-%20A%20Data%20Collector's%20Field%20Guide.pdf

Mathers, C., Oliva, M., & Laine, S. (2008*). Improving instruction through effective teacher evaluation: Options for States and Districts*. Washington, DC: National Comprehensive Centre for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/February2008Brief.pdf

Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data?. *Educational Evaluation and Policy Analysis*, *21*(1), 29-45.

McMillan, J. (1996). *Educational research fundamentals for the consumer*. New York: HarperCollins

Mears, C. (2012). In-depth interviews. In J. Arthur, M. Waring, R. Coe & L. Hedges (Eds.), *Research methods and methodologies in education*. London: SAGE.

Mertens, D. (2015). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (4^{th} ed.). London: SAGE.

Mertler, C. (1997). *Students as stakeholders in Teacher evaluation; teacher perceptions of a formative feedback model.* Paper presented at the annual meeting of the Mid-Western Educational Research Association. Chicago, IL.

Ministry of Education, Kuwait. (2011). *Performance evaluation*. Kuwait: Ministry of Education.

Ministry of Education, Kuwait. (2013a). *The organisational structure of the Ministry of Education in Kuwait*. Retrieved December 10, 2013, from http://www.moe.edu.kw/SitePages/home.aspx#

Ministry of Education, Kuwait. (2013b). *Education stages in the State of Kuwait*. Retrieved May 15, 2013, from http://www.moe.edu.kw/pages/misc/history/learning.htm

Ministry of Education, Kuwait. (2013c). *The framework for the Ministry of Education to develop Education*. Kuwait: Ministry of Education.

Montgomery, D. (1999). *Positive teacher appraisal through classroom observation*. London: David Fulton Publishers ltd.

Neal, D., & Schanzenbach, D. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, *92*(2), 263–283.

Netherlands Inspectorate of Education. (2009). *Risk-based inspection as of 2009 primary and secondary school*. Netherlands: Ministry of Education.

Netherlands Inspectorate of Education. (2012). *The Inspectorate of Education in the Netherlands*. Netherlands: Ministry of Education.

Nevo, D. (2001). School evaluation: Internal or External?. *Studies in Educational* Evaluation, *27*(2) 95-106.

Newby, P. (2014). *Research Methods for Education* (2nd ed.). New York: Routledge.

Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, *18*(23), 1-27.

Nicol, D. & Macfarlane-Dick, D (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218.

Nikolic, V. (2002). *Self-evaluation and improved teaching practice*. Washington, DC: ERIC Document Reproduction Services. Retrieved from http://files.eric.ed.gov/fulltext/ED468594.pdf

Nolan, J., & Hoover, L. (2008). *Teacher supervision and evaluation: Theory into practice* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625-632.

Norusis, M. (2008). *SPSS 16.0 Guide to Data Analysis*. Upper Saddle River, NJ : Prentice Hall

Nusche, D., Braun, H., Halasz, G., & Santiago, P. (2014). *OECD Reviews of Evaluation and Assessment in Education: Netherlands*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/edu/school/OECD-Evaluation-Assessment-Review-Netherlands.pdf.

Olatoye, R., & Aanu, E. (2011). Senior secondary school science teacher perception of using students to evaluate teaching effectiveness. *Journal of Emerging Trend in Education Research and Policy Studies, 2*(3), 164-170.

Oluwatayo, J. A. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research*, *2*(2), 391-400.

Ovando, M. (2001). Teachers' perceptions of a learner-centered teacher evaluation system. *Journal of Personnel Evaluation in Education, 15*(3), 213-231.

Partee, G. (2012). *Using multiple evaluation measures to improve teacher effectiveness: State strategies from round 2 of No Child Left Behind Act waivers*. Washington, DC: Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/2012/12/MultipleMeasures-2.pdf

Patton, M. (2002). *Qualitative evaluation and research method* (3rd ed.). London: SAGE

Peterson, K., Wahlquist, C. & Bone, K. (2000) Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, *14*(2) 135-153.

Peterson, P., & Comeaux, M. (1990). Evaluating the systems: Teachers' perspectives on teacher evaluation. *Educational Evaluation and Policy Analysis, 12*(1), 3-24.

Phillips, V., & Weingarten, R. (2013). The professional educator: Six steps to effective teacher development and evaluation. *American Educator*, *37*(2), 36-37.

Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology, 30*(4), 59-64.

Rockoff, J., & Turner, L. (2008). Short run impacts of accountability on school quality. *National Bureau of Economic Research Working Paper*, No. 14564.

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education, 28*(2), 147-164.

Ryan, K., Chandler, M., & Samuels, M. (2007). What should school-based evaluation look like?. *Studies in Educational Evaluation*, *33*(3), 197-212.

Sabti, A. (2010). Criteria and procedures of teacher evaluation. *Proceedings of the Educational Conference of Kuwait Teachers Society, Kuwait, 39*(1582).

Sadler, R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*(5), 535-550.

Salih, A. (2013). Peer evaluation of teaching or fear evaluation: In search of compatibility. *Higher Education Studies, 3*(2),102-114.

Santiago, P., & Benavides, F. (2009, December). *Teacher evaluation: A conceptual framework and examples of country practices*. Paper presented at OECD-Mexico Workshop Towards a Teacher Evaluation Framework: International Practices, Criteria and Mechanisms, Mexico. Retrieved from http://www.oecd.org/edu/school/44568106.pdf

Schafer, W., Lissitz, R., Zhu, X., Zhang, Y., Hou, X., & Li, Y. (2012). Evaluating teachers and schools using student growth models. *Practical Assessment, Research & Evaluation*, *17*(17), 1-21.

Scheeler, M., Ruhl, K. & McAfee, J. (2004). Providing performance feedback to teacher: A Review. *Teacher Education and Special Education, 27*(4), 396-407.

Scheerens, J., Ehren, M., Sleegers, P., & de Leeuw, R. (2012). *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country background report for The Netherlands*. Paris: OECD Publishing. Retrieved from http://www.oecd.org/edu/school/NLD_CBR_Evaluation_and_Assessment.pdf

Silverman, D. (2010) *Doing qualitative research: Practical handbook* (3rd ed.). London: SAGE.

Smith, M., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational measurement: Issues and practice*, *10*(4), 7-11.

Smith, R. (2001). Formative evaluation and the scholarship of teaching and learning. *New Directions for Teaching and Learning*, winter *2001*(88), 51-62.

Smith, S., & Mickelson, R. (2000). All that glitters is not gold: School reform in Charlotte-Mecklenburg. *Educational Evaluation and Policy Analysis*, *22*(2), 101-127.

State Collaborative on Reforming Education, SCORE (2012). *Supporting effective teaching in Tennessee: Listening and gathering feedback on Tennessee's teacher evaluations*. Retrieved from http://tnscore.org/scorereports/supporting-effective-teaching-in-tennessee-listening-and-gathering-feedback-on-tennessees-teacher-evaluations/

Stecher, B. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In Stecher, B., Hamilton, L., & Klein, S. (Eds.), *Making Sense of Test-Based Accountability in Education*. California: RAND.

Stronge, J. (2006). Teacher evaluation and school improvement: Improving the educational landscape. In J. Stronge (Ed.), *Evaluating teaching* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education

Tashakkori, A., & Creswell, J. (2007). Editorial: The new era of mixed methods. *Journal of Mixed Methods Research*, 1, 3-7.

Taylor, E., & Tyler, J. (2011). *The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers*. Washington DC: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w16877.pdf

Thomas, G. (2011). *How to do your case study: A guide for students and researchers*. London: SAGE.

Thorndike, R., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). London: Pearson Education.

Tuytens, M., & Devos, G. (2012). Importance of system and leadership in performance appraisal. *Personnel Review, 41*(6) 756-776.

Tymms, P (2012). Questionnaire. In J. Arthur, M. Waring, R. Coe & L. Hedges. (Eds.), *Research methods and methodologies in education*. London: SAGE.

Vanhoof, J., & Van Petegem, P. (2007). Matching internal and external evaluation in an era of accountability and school development: Lessons from a Flemish perspective. *Studies in Educational Evaluation*, *33*(2), 101-119.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project. Retrieved from http://files.eric.ed.gov/fulltext/ED515659.pdf

Westhuizen, G., & Smith, K. (2000). Teachers' portfolio reflections: A Comparative analysis. *Teacher Development, 4*(3), 339-352.

Whitehurst, G., Chingos, M., & Lindquist, K. (2014). *Evaluating teachers with classroom observations: Lessons learned in four Districts*. Washington, DC: Brown Centre on Education Policy. Retrieved from http://www.brookings.edu/~/media/research/files/reports/2014/05/13-teacher-evaluation/evaluating-teachers-with-classroom-observations.pdf

Wolf, I. & Verkroost, J. (2011). *Evaluation of theory and practice of risk-based school inspections in the Netherlands*. Utrecht: The Netherlands Inspectorate of Education. Retrieved from http://www.sici-inspectorates.eu/getattachment/ee4990f8-363d-4437-8548-f0c0d8756499

Wong, M., Cook, T., & Steiner, P. (2009). *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series.* Evanston: Northwestern University Institute for Policy Research.

Yong, A., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94.

Zhang, X. (2008). *The role of teacher appraisal in teacher professional development: A case study in schools in Shanghai* (Doctoral dissertation, The University of Hong Kong, Pokfulam, Hong Kong). Retrieved from http://hub.hku.hk/bitstream/10722/51240/3/FullText.pdf?accept=1

Appendices

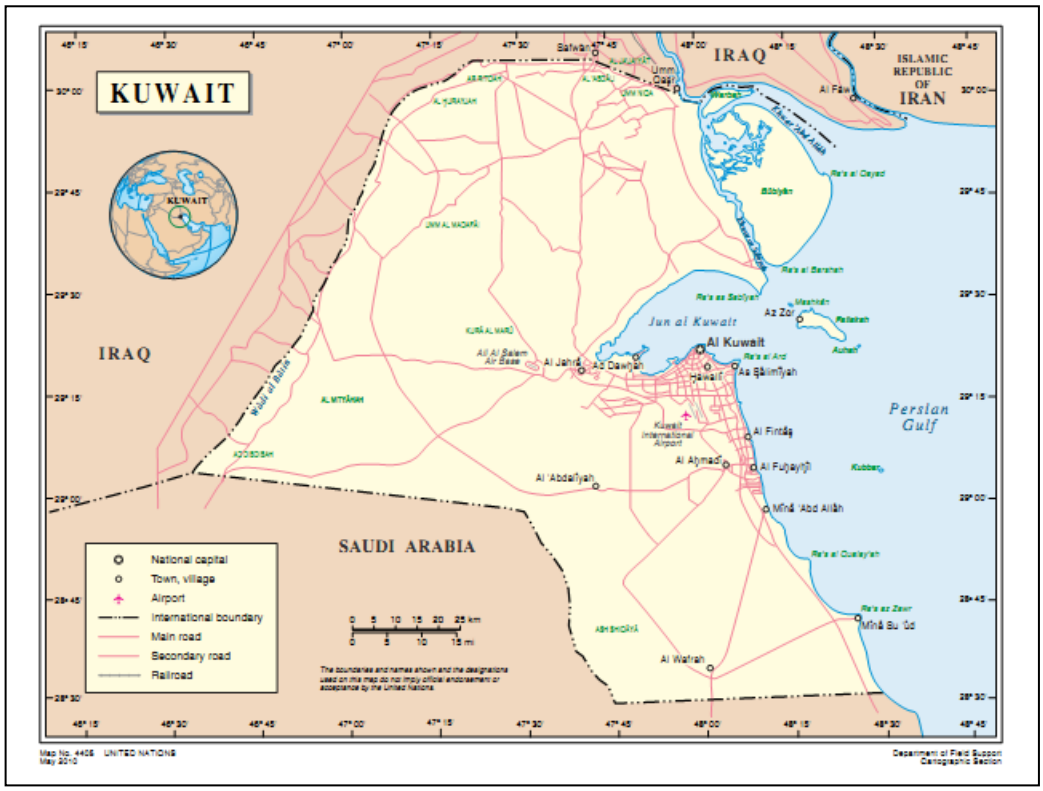
Appendix (1): The Context of Kuwait

Introduction:

Kuwait became an independent country and a member of the League of Arab States in 1961. In 1963 it became a member of the United Nations (Kuwait National Assembly, 2013). Kuwait is located in the Middle East in the north-western corner of the Arabian Gulf. It is bordered to the south and south-west by the Kingdom of Saudi Arabia, to the north and north-west by the Republic of Iraq, and to the east by the Arabian Gulf. The official language of Kuwait is Arabic, and the official religion for the state is Islam (Official E-Portal for the State of Kuwait, 2014).

The total area of Kuwait is 17,818 square kilometres. The population in 2013 was 3,448,139 with 1,159,787 of the population having Kuwaiti Nationality. The others are non-Kuwaitis who work or live in Kuwait (Kuwait Central Statistical Bureau [KCSB], 2015). The population is concentrated in Kuwait city and its surrounding areas, close to the coast of the Arabian Gulf. See Figure (1) (KCBS, 2013/2014).

Figure (1) Kuwait Map.



According to the Kuwaiti Constitution that was established in 1962, the system of government in Kuwait is democratic. Article 6 states, "Kuwait's system of government is democratic; sovereignty is vested in the nation as the source of all authority; and the exercise of that sovereignty shall be as set out in this Constitution" (Kuwait Constitution, 2008, p.4).

History of Education in Kuwait:

As in other Islamic countries, in Kuwait prior to 1911, education was informal, and the 'Mosque' was the place for teaching people about Islam from The Holy Quran; teaching Hadiths - 'Sayings of the Prophet Muhammad'. Thereafter, a new education system entitled 'Alkatateeb' was established. This is where some people who have received a good education open their houses to teach people subjects such as Islamic studies, Mathematics and Arabic. This can be free or can be paid for. This system separated males from females, with 'Almullah' teaching male students and 'Almuttawwa' teaching female students. They were the teachers in those days (Kuwait News Agency [KUNA], 2002; Centre for Research and Studies on Kuwait [CRSK], 2002).

On the 22nd December 1911, formal education was established with the opening of the first school in Kuwait which was named 'Al-Mubarakiya', and was for male students. It was built and funded by the denotations of the Kuwaiti people (KUNA, 2008). After that, the second boy's school, 'Al-Ahmadiyya', was established in 1921. In 1936, the first school for female students - 'Al-Wusta' - was established (CRSK, 2002).

As Kuwait developed, over the years the number of schools increased, and it was necessary to establish a council to organise education in Kuwait. Therefore, in 1936, the Council of Knowledge was established to manage and fund schools in Kuwait. In 1962 the Council of Knowledge was changed to the Ministry of Education (KUNA, 2002; Alhatem, 1980).

Education in Kuwait at the present time:

Nowadays, the education of the Kuwaiti people is the full responsibility of the Kuwaiti government. In the Kuwaiti Constitution of 1962, Article (40) states:

> Education is a right for Kuwaitis, guaranteed by the State in accordance with law and within the limits of public policy and morals. Education in its preliminary stages shall be compulsory and free in accordance with the law. The law shall lay down the necessary plan to eliminate illiteracy. The State shall devote particular care to the physical, moral and mental development of youth (Kuwait Constitution, 2008).

As a result of this constitutional provision, education was freely extended to all levels and every type of education, as is now the case in Kuwait, with the exception of private schools and private universities.

Ministry of Education:

The Kuwaiti government provides free education for citizens and controls education through the Ministry of Education (MOE). The organisational structure of the MOE is as shown in Figure (2):

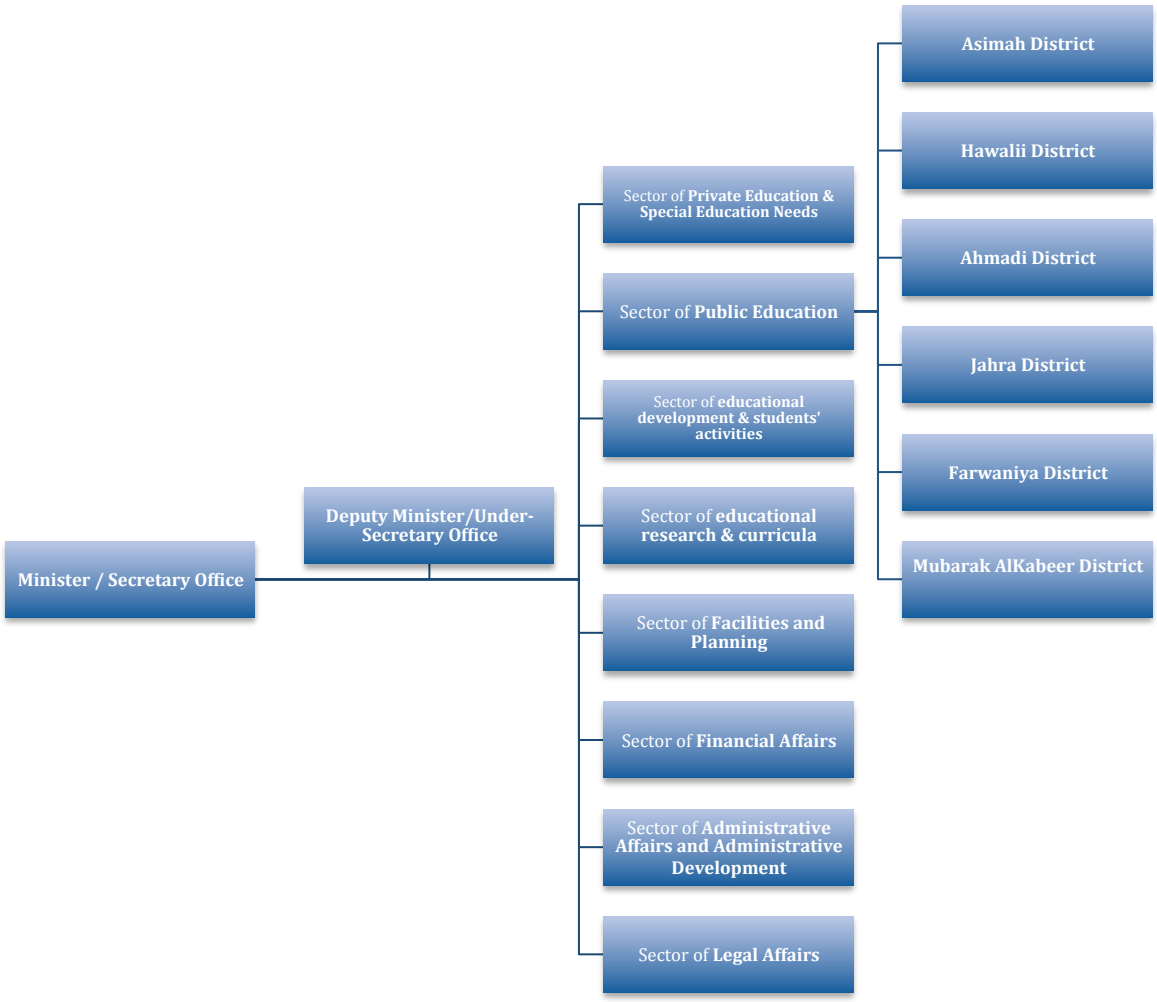


Figure (2) The organisational structure of the MOE (MOE, 2013a).

Figure (2) demonstrates that the education system is controlled by the MOE, in that the Minister implements regulations and decisions, and ensures follow-up on the part of his or her Deputy Minister. The Deputy Minister or Under-Secretary works with his or her team of assistants and is responsible for several sectors under the auspices of the MOE (2013a). These are as follows:

A) The private education and special education needs sector: all private and special education needs schools are under the MOE's responsibility in terms of working to organise and follow-up these types of schools in line with the regulations that have been introduced for this type of education.

B) Public education sector: This sector relates to all public schools. The Deputy Minister or Under-Secretary is responsible for the work done in schools, students, teachers, heads of departments, head teachers in public schools and the inspectorate department. The Assistant Under-Secretary works with his or her team of managers over six districts. These are Asimah, Hawalli, Ahmadi, Jahra, Farwaniya and Mubarak AlKabeer.

C) Educational development and students' activities sector: this is responsible for libraries, technology-enhanced learning, students' clubs, psychological and social counseling.

D) Educational research and curricular sector: this sector is responsible for conducting research and studies for the MOE, and for the design of national curricula for each level of school.

E) Facilities and planning sector: this sector is responsible for the construction of schools and facilities and other planning, such as the planning and design of new school buildings.

F) Financial affairs sector: this sector is responsible for all financial affairs concerning the funding of schools, and the activities and salaries for all MOE employees.

G) Administrative affairs and administrative development sector: this sector is responsible for human resource management, such as coordination and appointments, promotions, distributing teachers to districts and schools, etc. As for administrative development, this includes scholarships for employees, and MOE training courses.

H) Legal affairs sector: this sector is, for example, concerned with the design and explaining of regulations with regard to organising work, determining and giving authorisation for work, investigating issues, and modifying laws if necessary.

Education system:

The school system in Kuwait is divided into four levels: kindergarten, primary school, middle school and high school. The schools are separate for girls and boys. This section provides an explanation of the upper three levels of school - primary, middle and high.

According to the MOE (2013b) in 2004/2005, it changed the number of stages in school in Kuwait. These became 5 stages in primary school, 4 stages in middle school and 3 stages in high school. For each schooling stage, students should study for one year and achieve 'pass' to move to the next stage. The education system is outlined as follows:

Primary school is for students between six and ten years of age. It is compulsory for children to attend at this level with sanctions for parents who do not abide with the regulation. Students are taught basic skills and some specific subjects as part of the National Curriculum: Arabic, English, Mathematics, Islamic Studies, Science, Social Studies, and Computer Science. There are also some subjects that are taught such as Art and Sports Education with regard to which the students are not examined.

Middle school is for students between eleven and fourteen years of age, with four stages – the sixth, seventh, eighth and ninth stages. It also is compulsory, and students are examined through the use of four test periods in the school year. In terms of subjects, these are the same as in primary school, but with more in-depth information and on a higher level.

In high school, students are aged between fifteen and seventeen/eighteen years, with three stages. In the first stage, which is called the tenth stage, students are taught Social studies, Constitution, Arabic, English, Islamic studies, students are taught Physics, Chemistry, Biology and Mathematics. There are also some subjects that are taught but on which the students are not examined, such as Art, Technical Studies and Sports Education.

In the eleventh stage, Subjects are divided into two sections: Arts and Science. Within Arts, students are taught Geography, History, Constitution, Sociology, Philosophy and French. Within Science, students are taught Physics, Chemistry, Biology and Mathematics. Students select the Arts or Science stream. However, students in both streams continue studying English, Arabic and Islamic studies as compulsory modules. In the twelfth stage,

students will complete their stream to finish high school.

In terms of curricula and testing, Kuwait has a national curriculum for every type of school. The testing depends on the school level. In primary and middle schools, the exams are designed separately by each educational district, but are arranged in the same time period in all six educational districts. In high school, examinations are unified (national achievement standardised tests) in all six educational districts simultaneously. Tests are conducted in four periods throughout the school year for all levels.

Every type of school has a department for each taught subject. In each department, there are approximately six teachers depending on the needs of the department; for example, the number of classes in school and the number of weekly lessons for each subject. There are one or two heads of department: their role is to organise school activities, continuously supervise the teachers, evaluate teachers, distribute classrooms to the teachers in their department, as well as teaching one or two lessons for one class.

Each school should also have a head teacher and two or three assistant heads. They are involved in the management of the school, and are responsible to the educational district for running the school. Their role is to organise and manage the school, to monitor the attendance of teachers and other staff, and provide for the needs of the school. The head of the school also has to participate in evaluating teachers on an annual basis.

With regard to inspectors, they work in the Department of Inspection of the MOE. Every subject that is taught in Kuwait has a main department of inspection (Arabic, English, Mathematics, Islamic Studies, Science, Social Studies, French, Computer Science, Arts, Sports Education). Each department is divided into six sub-departments for each of the six districts. For example, the Arabic inspection department is divided into sub-departments dealing with the Ahmadi, Hawalli, Asimah, Jahra, Farwaniya and Mubarak AlKabeer districts. Each sub-department has inspectors who are responsible for all types of schools in their district. They are responsible for designing and improving the national curricula, evaluating teachers, and designing examinations for schools.

Teachers:

According to the Kuwait Teachers Society [KTS] (2012a) there are regulations concerning the organisation of teaching jobs and of teachers in public schools:

A) 'Teacher' candidates should have a Bachelor of Education or higher education degree, or any bachelor degree that is related to the subjects that are taught in schools, e.g. Chemistry, Mathematics etc.
B) Candidates should be nominated by the Civil Service Commission [CSC] to the 'appointments section' of the administrative affairs sector in the MOE. Therefore, candidates should register with the CSC when they are seeking a teaching post. In some cases, there are conditions set by the administrative affairs sector, such as passing a job interview if this is deemed necessary.
C) Candidates should have Kuwaiti nationality. However, if there are no Kuwaiti applicants, the MOE can appoint from other Arab countries or non-Arab countries. However, priority is given to Arabs.
D) Other conditions, such as good conduct and health status etc.

In term of candidates in private schools, the same conditions as above should apply with the exception of condition (B) and (C). Teachers may be appointed by the school management with the consent of the private education sector of the MOE.

Promotion regulations for teachers are as mentioned in the MOE (2011b). A teacher in public schools can become a head of department. After becoming head of department, there is a choice of two directions in terms of promotion; either to be an inspector or an assistant head teacher of a school. An assistant head teacher will be in a position to gain experience and may later become a head teacher of a school and to be promoted to an educational observer. If a head of department chooses to be an inspector in his or her subject, there is a possibility to become a head inspector and to be promoted to general inspector of the subject. Figure (3) shows the organisational structure with regard to promotion in the teaching profession.

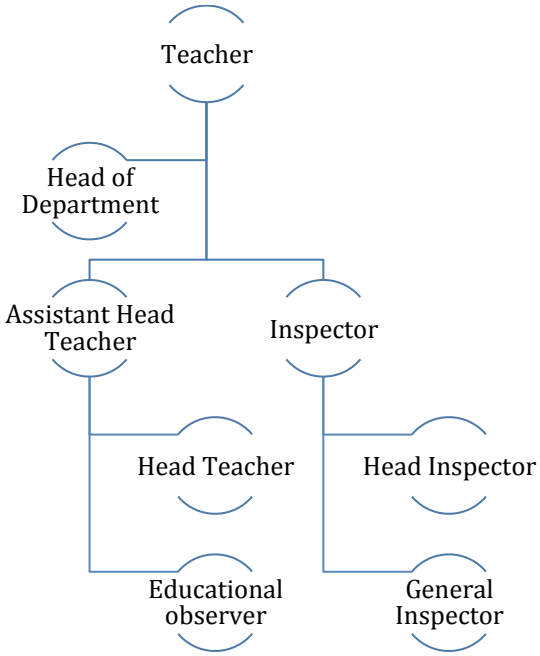


Figure (3) The promotional structure for teachers in the MOE.

Teacher Evaluation:

In this section, the previous and current teacher evaluation systems in Kuwait are presented. Where there are two systems of teacher evaluation, the previous one which no longer exists is 461/1993, and the current system is 36/2006.

Teacher Evaluation 461/1993 (No longer exists):

According to ALkhayat and Dhiab (1996) and Alhamdan (1998), this system was established in September 1993. Individual teachers were evaluated by an inspector and head teacher on an annual basis. Teacher evaluation was divided into three parts:

1) A follow-up card provides a daily and weekly record. It consists of a teacher's demographic data, his/her commitment to work, attendance and absence levels, training programmes undertaken, visitors' observation checklist and teachers' activities.
2) The mid-year report of teachers' performance in the first semester. This consists of self-evaluation by the teacher and the opinion of an inspector and his/her head teacher. Teachers are informed of the report and their score.

3) The final individual teacher performance report for the school year. Teachers are not informed about the final evaluation report, as they must be kept confidential,

except in the case of a weak teacher who is informed of the contents of his or her evaluation report. The score for teacher evaluation in this system has three levels: outstanding indicated by a score of 90-100; good for a score between 60-89; and weak for a 59 or lower score. The criteria for the final report are as shown in Table (1).

No.	Criteria of teacher evaluation
1	Teacher takes responsibility.
2	Good conduct.
3	Cooperation with school staff and colleagues.
4	Teachers' commitment to attendance and absence.
5	Implementation of policies, decisions and regulations with regard to work.
6	Teachers' commitment to the ethics of the profession.
7	Adoption of new developments in education.
8	Preservation of public property.
9	Time management with regard to his or her work.
10	Achievement of educational targets.
11	Teacher knowledge and mastery of the content of the subject that is taught.
12	Lesson planning.
13	Classroom management.
14	Pedagogies.
15	Classroom activities and using tools to enhance learning.
16	Teacher relationship with students and communication.
17	Follow up with students and ability to overcome students' weaknesses or problems.
18	Participating in and preparing school activities.
19	Students' activities and assessment of students.

Table (1) The teacher evaluation criteria for 461/1993.

Teacher Evaluation 36/2006 (The current system):

The Civil Service Commission (CSC) has changed employee performance evaluation by Regulation Number 36/2006. Based on this decree, the MOE changed the teacher evaluation system in 2006, and continues to use if to the present day (KTS, 2010)

The purposes of teacher evaluation are: to determine individual teacher performance during the school year accurately and objectively, and to develop teachers' performance. Teacher evaluation is also used to make decisions about promotion (MOE, 2011a; KTS, 2010) and about increases in teacher's salary and annual bonuses as noted in the Salary Scale of Teachers Number 48/2011 (KTS, 2012b), or make decisions about either dismissal or transfer to the administrative staff as sanctioned (KTS, 2010).

In this system, every teacher is evaluated by three evaluators: the head of department, an inspector and the head teacher. This system is based on a written evaluation report by each evaluator at the end of the school year. The final teacher evaluation report is entered in the CSE's online portal by the head teacher, after agreement among the three evaluators. In addition, the head teacher has to print the final individual teacher's evaluation. This must then be signed by the evaluators and sent by the educational observers in the educational district for insertion into the teachers' record in the MOE (MOE, 2011a; KTS, 2010).

According to KTS (2010) a final teacher evaluation consists of formal criteria using the following scales: the efficiency of individual performance, the efficiency of collective performance and the efficiency of personality. Table (2) shows the criteria for each scale and score for each criterion. Teachers are not informed about the final report of their evaluation, except in the case of a weak teacher who is told about his or her evaluation result after it has been signed.

Scale	The Criteria	Rate for each criteria
The efficiency of individual performance	1. Teacher's commitment to the work, including attendance and absences. 2. Completion and precision of work such as preparing lessons and commitment to teaching the content of the curriculum on time, doing what require of them such as in terms of school activities, assessing students 3. Good conduct and taking responsibility. 4. Teachers' commitment to the ethics of the profession, including teacher behaviour. 5. Preservation of public property. 6. Teacher interaction and dealing with families and students. 7. Deep understanding and mastery of the subject content that is taught and of skills which are related to teaching 8. Providing suggestions and studies, whether written or oral, which contribute to development work. 9. Teachers' commitment to the decisions and administrative instructions, regulations and decrees of the MOE.	The rating for each criterion is from 3 as a minimum to 10 as a maximum.
The efficiency of personality	1. Teacher's appearance, such as "dress appropriately; national dress for male teachers with the exception of physical and drawing teachers and non-Kuwaiti teachers" and teacher behaviour. 2. The extent to which teachers accept criticism and suggestions with regard to their work. 3. Self-development in teaching and learning.	The rating for each criterion is from 3 as a minimum to 10 as a maximum.
The efficiency of collective performance	1. The extent to which teachers cooperate with colleagues and school staff. 2. The extent to which teachers are interested in sharing experiences with others. 3. Deep understanding and mastery of the educational targets of the MOE and how well they achieve those targets.	The rating for each criterion is from 3 as a minimum to 10 as a maximum.

Table (2) The criteria of teacher evaluation system 36/2006 in Kuwait (KTS, 2010)

In this system the teacher evaluation score is on four levels: outstanding is 90 and more, very good is 75-89, good is 55-74 and weak lower than 55. See Table (3)

No.	Teacher Evaluation Score		
	Grade	From	To
1	Outstanding	90	100
2	Very Good	75	89
3	Good	55	74
4	Weak	0	55

Table (3) Score for the 36/2006 teacher evaluation system

With regard to specific tools that should be used to collect evidence about teacher performance during the school year such as classroom observations, student achievement, etc. are not mentioned in particular in terms of the policy of teacher evaluation. However, the teacher evaluation policy stipulates that evaluators have to determine teacher's performance both inside and outside the classroom, and use a standardised checklist for determining teachers' performance inside the classroom.

Reference

Alhatem, A. (1980) *From here established Kuwait*. Kuwait: Dar Al-Qabs

Alhamdan, J. (1998). Teacher evaluation in the State of Kuwait. *Educational Journal: Kuwait University, 12*(47), 289-312.

Alkhayat, A. & Dhiab, A. (1996). Teacher evaluation system in the Ministry of Education in Kuwait: Evaluation study. *Educational Journal: Kuwait University, 10*(38), 27-78.

Centre for Research and Studies on Kuwait (2002). *The history of the State of Kuwait*. Kuwait: CRSK.

Kuwait Central Statistical Bureau. (2013/2014). *Annual statistical abstract*. Kuwait: Kuwait Government.

Kuwait Central Statistical Bureau. (2015). *Annual statistical abstract*. Kuwait: Kuwait Government.

Kuwait Constitution. (2008). *The Constitution of the State of Kuwait*. Kuwait: Ministry of Information.

Kuwait National Assembly. (2013). *Democratic life in Kuwait*. Retrieved September 10, 2013, from http://www.kna.kw/clt/run.asp?id=161#sthash.HOOC6MDl.dpbs.

Kuwait News Agency. (2008). *It is happened in Kuwait*. Kuwait: KUNA.

Kuwait News Agency. (2002). *Education in Kuwait from Alkatateeb to universities, institutes and centres: Historical documentary*. Kuwait: KUNA

Kuwait Teacher Society. (2010). Guidance and the policy of teacher evaluation: Number 36/2006. *Kuwait Teacher society Journal*, *June* (1586).

Kuwait Teachers Society. (2012a). *Legal guide for teachers* (2nd ed.). Kuwait: KTS

Kuwait Teacher Society. (2012b). *Cadre of Teachers No. 28/2011*. Kuwait: Kuwait Teacher Society.

Ministry of Education, Kuwait. (1993). *Ministerial Decree, No 146/1993*. Kuwait: MOE

Ministry of Education, Kuwait. (2011a). *Performance Evaluation*. Kuwait: Ministry of Education.

Ministry of Education, Kuwait. (2011b). Promotions for teaching staff in all school levels 2011/2012. Kuwait: MOE.

Ministry of Education, Kuwait. (2013a). *The Organisational Structure of the Ministry of Education in Kuwait*. Retrieved December 10, 2013, from http://www.moe.edu.kw/SitePages/home.aspx#

Ministry of Education, Kuwait. (2013b). *Education Stages in the State of Kuwait*. Retrieved May 15, 2013, from http://www.moe.edu.kw/pages/misc/history/learning.htm

Official E-Portal for the State of Kuwait. (2014). *About Kuwait*. Retrieved August 13, 2014, from http://www.e.gov.kw/sites/kgoarabic/portal/Pages/Visitors/AboutKuwait.aspx.

Appendix (2): The Complete of the Questionnaire [English]

Dear Teacher,

I am a PhD student in the School of Education at Durham University in the UK. I have conducted this research to analyse the current teacher evaluation system in Kuwait from teachers' and evaluators' perspectives. I would be grateful if you help me by answering this questionnaire.

Your answer will be useful for teachers and their evaluation system because it will allow you to add significant information about teacher evaluation system.

Finally, I would like to inform you that all the information collected will only be used for academic purposes; the research will not cause you any harm as no personal information will be revealed at any time.

The Researcher

Talal S. Almutairi
PhD Student
School of Education
Durham University
England, UK
Email: Talal.Almutairi@Durham.ac.uk

Personal information:

Gender: ( ) Male ( ) Female

Position: ( ) Teacher ( ) Head-Department

Experience (Years in Teaching): ……

Subject Area:……..

Educational District:……

Section One: (The Purposes of Teacher Evaluation)

The teacher evaluation system has several purposes as listed below. In your view, **how are these used in practice /and how important are these from your perspective.** Rate each purpose:

e.g.:

Statement	**How often are these used**						**How important are these to be used**				
	Never	Seldom	Sometimes	Often	Always		Not Important at all	Unimportant	Neither Important nor unimportant	Important	Very Important
Promoting professional development of teachers			**X**							**X**	

Statement	**How often are these used**						**How important are these**				
	Never	Seldom	Sometimes	Often	Always		Not Important at all	Unimportant	Neither Important nor unimportant	Important	Very Important
1) Promoting professional development of teachers											
2) Determining the teacher's performance											
3) Supporting decision-makers to make decisions about teachers that are related to sanctions or rewards											

Section Two: (The tools of teacher evaluation)

When you have been evaluated, **(A)** to what degree the tools are used ‘and **(B)** to what degree the tools should be used.

Statement	Is used						Should be used				
	Never	Seldom	Sometimes	Often	Always		Never	Seldom	Sometimes	Often	Always
4) You have been evaluated by classroom observation											
5) Evaluators have been used student’s achievements as a data resource for your evaluation											
6) Your ‘self-evaluation’ is used as data for teacher evaluation											
7) You have been evaluated by your colleagues in order to use for formative purpose “to provide feedback to teacher, not use for judging [score]”											
8) You have been evaluated by students, either by survey or a focus group interview											
9) Evaluators have used portfolios that includes different types of teachers’ work in the school year (for example; school activities, attending conference, workshop) as data resource for your evaluation											

Section Three: (The involvement of evaluators)

When you have been evaluated, **to what degree has the following happened with each of the listed evaluator, and how do you rate the value each of these**?

Statement (Head Teacher)	**Has happened**					**Rating the value**				
	Never	Seldom	Sometimes	Often	Always	Poor	Fair	Good	Very Good	Excellent
10) You have had a discussion before classroom observation										
11) You have had a discussion after classroom observation										
12) You have received written feedback										

Statement (Inspector)	**Has happened**					**Rating the value**				
	Never	Seldom	Sometimes	Often	Always	Poor	Fair	Good	Very Good	Excellent
13) You have had a discussion before classroom observation										
14) You have had a discussion after classroom observation										
15) You have received written feedback										

Statement (Head Of Department) **For Teacher Only**	**Has happened**					**Rating the value**				
	Never	Seldom	Sometimes	Often	Always	Poor	Fair	Good	Very Good	Excellent
16) You have had a discussion before classroom observation										
17) You have had a discussion after classroom observation										
18) You have received written feedback										

Section Four:

What support have you personally had from taking part in teacher evaluation?

Statement	Strongly disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
19) It has improved the deep understanding of content that you teach					
20) It has assisted you with better use of pedagogies					
21) It has given you much clearer understanding of lesson planning					
22) It has given you much clearer understanding of what constitutes effective teaching					
23) It has revealed the weaknesses of your performance					
24) It has played a significant role in determining the strengths of your performance					
25) It has affected your organisation of activities in the classroom					
26) It has affected your dealing with students' discipline and behaviour problems					
27) It has affected on your ability to motivate students in terms of their learning					
28) It has affected your ability to deal with individual differences between students					
29) It has affected your continuous assessment of students' learning					
30) It has affected your ability to provide students with effective feedback					
31) It has affected your rewards in terms of annual bonus or salary increase					
32) It has impacted you in terms of your promotions					

Note: If you would like to participate in focus group in order to present the proposed alternative system and to obtain perspective of teachers toward a new system for teacher evaluation. Please write your email address or your name:…….. …….. the researcher will contact you soon, or please send me message on +965 50660094 or email on Talal.Almutairi@durham.ac.uk, include you name and your school.

Appendix (3): The Complete of Questionnaire [Arabic]

عزيزي المعلم

تحية طيبة وبعد,,,

أود إفادتك بأنني أحد طلاب الدكتوراه في جامعة درم بالمملكة المتحدة / كلية التربية في تخصص القياس والتقويم التربوي. وأقوم حالياً بدراسة حول نظام تقويم المعلم في دولة الكويت. سأكون ممتناً إذا تفضلت بمساعدتي من خلال هذا الاستبيان.

حقيقة أنا على يقين أن تنفيذ هذه الاستبانة معك سيضيف للمعلم ولتقويم المعلم الكثير, وبالتالي سيصبح المعلم المساهم الأول في التحسين والتطوير لنظام تقويمه.

أخيراً, أود أن أؤكد لك أنه سيتم التعامل مع المعلومات وآرائك الشخصية بشكل سري لأغراض البحث العلمي وأهداف هذه الدراسة فقط, وتفضلوا بقبول خالص الشكر والتقدير ,,,

الباحث

طلال سعد المطيري

القياس والتقويم

برنامج الدكتوراه / كلية التربية

جامعة درم - المملكة المتحدة

Email: Talal.Almutairi@durham.ac.uk

المعلومات الشخصية

الجنس: ( ) ذكر ( ) أنثى

الوظيفة: ( ) معلم ( ) رئيس قسم

الخبرة (سنوات العمل في التدريس): .....

المادة العلمية: .....

المنطقة التعليمية: .....

القسم الأول: أهداف تقويم المعلم

نظام تقويم المعلم له عدة أهداف كما هو مبيّن أدناه، **إلى أي مدى يستخدم تقويم المعلم في الكويت لتحقيق الأهداف المذكورة، وما مدى أهمية** هذه الأهداف للمعلم من وجهة نظرك

ضع معياراً لكل غرض كما هو موضح بالمثال:

مثال للإجابة:

البيان	إلى أي مدى يستخدم تقويم المعلم لتحقيق هذه الأهداف في الواقع العملي						مامدى أهمية هذه الأهداف للمعلم من وجهة نظرك					
	أبداً	نادراً	أحياناً	غالباً	دائماً		غير مهم إطلاقاً	غير مهم	لا أستطيع أن أحدد	مهم	مهم جداً
لتعزيز التطوير والتحسين المهني للمعلم			X						X		

البيان	إلى أي مدى يستخدم تقويم المعلم لتحقيق هذه الأهداف في الواقع العملي						مامدى أهمية هذه الأهداف للمعلم من وجهة نظرك					
	أبداً	نادراً	أحياناً	غالباً	دائماً		غير مهم إطلاقاً	غير مهم	لا أستطيع أن أحدد	مهم	مهم جداً
١) تعزيز التطوير المهني للمعلم.											
٢) تحديد مستوى أداء المعلم.											
٣) لأخذ قرارات تجاه المعلم، ذات صلة بالثواب أوالعقاب. [مثل: الترقية الوظيفية، الحوافز، الفصل]											

القسم الثاني: أدوات التقويم

عندما تقوّم كمعلم من خلال نظام تقويم المعلّم المتّبع في دولة الكويت ، إلى أي درجة قد استخدمت الأدوات التالية **أثناء** وإلى أي درجة تقترح استخدام هذه الأدوات في تقويمك ؟ **/تقويمك**

البيان	إلى أي درجة إستخدمت هذه الأدوات في تقويمك						إلى اي درجة تقترح استخدام هذه الأدوات في تقويمك				
	لا أعرف	بدرجة قليلة	بدرجة متوسطة	بدرجة كبيرة	بدرجة كبيرة جدا		لا أعرف	بدرجة قليلة	بدرجة متوسطة	بدرجة كبيرة	بدرجة كبيرة جدا
٤) تم تقويمك من خلال الملاحظة الصفية من قبل المقوّمين.											
٥) تم استخدام التحصيل العلمي للطلبة كمصدر لتقويمك من قبل المقوّمين.											
٦) تم استخدام تقويمك الذاتي كمصدر معلومات لتقويمك من قبل المقوّمين.											
٧) تم تقويم أدائك من قبل الزملاء، ويتم استخدام ذلك بهدف التحسين والتطوير، وليس لاتخاذ قرار متعلق بالأداء.											
٨) تم تقويمك بواسطة الطلبة سواء من خلال مثال: توزيع الاستبيان [الاستبيان أو مقابلة الطلبة على الطلبة من قبل التوجية او الادارة لأخذ الأراء]حول المعلم											
٩) استخدم المقيمون الملف الإنجازي الخاص بك، الذي يتضمن بعض أعمالك [مثل: الأنشطة المدرسية، حضور المؤتمرات، الورش] خلال العام الدراسي كمصدرمعلومات لتقويمك											

القسم الثالث:

عند تقويمك، **إلى أي درجة حدث الآتي مع جميع المقوّمين (مدير المدرسة، الموجه الفني، رئيس القسم). وماهو تقييمك لجودة كل واحدةٍ منها ان حدث ذلك .**

تقويمك من خلال مدير المدرسة	قد حدث مع المقيم بالفعل						مدى استفادتك ممّا حدث				
	أبداً	نادراً	أحياناً	غالباً	دائماً		ضعيف	مقبول	جيد	جيد جداً	ممتاز
١٠) المناقشة مع المقوّم قبل زيارته للفصل.											
١١) المناقشة مع المقوّم بعد زيارته لك الفصل.											
١٢) الحصول على تغذية راجعة مكتوبة.											

تقويمك من خلال الموجه الفني	قد حدث مع المقيم بالفعل						مدى استفادتك ممّا حدث				
	أبداً	نادراً	أحياناً	غالباً	دائماً		ضعيف	مقبول	جيد	جيد جداً	ممتاز
١٣) المناقشة مع المقوّم قبل زيارته للفصل.											
١٤) المناقشة مع المقوّم بعد زيارته لك الفصل.											
١٥) الحصول على تغذية راجعة مكتوبة.											

تقويمك من خلال رئيس القسم **[للمعلمين فقط]**	قد حدث مع المقيم بالفعل						مدى استفادتك ممّا حدث				
	أبداً	نادراً	أحياناً	غالباً	دائماً		ضعيف	مقبول	جيد	جيد جداً	ممتاز
١٦) المناقشة مع المقوّم قبل زيارته للفصل.											
١٧) المناقشة مع المقوّم بعد زيارته لك الفصل.											
١٨) الحصول على تغذية راجعة مكتوبة.											

القسم الرابع:

إلى اي مدى من وجهة نظرك كان نظام المعلم داعم لك شخصياً،(مالذي نتجت عن تقويمك من خلال نظام تقويم المعلم في دولة الكويت) ؟

البيان	لا أوافق بشدة	لا أوافق	محايد	أوافق	أوافق بشدة
١٩) أدّى إلى تحسين إلمامك بالمادة العلمية.					
٢٠) أسهم في مساعدتك في استخدام أفضل طرق التدريس					
٢١) زودك بالفهم الصحيح لتحضير وإعداد الدروس.					
٢٢) زودك بالفهم الدقيق لماهية التدريس الفعّال.					
٢٣) كشف لك نقاط الضعف في أدائك.					
٢٤) أظهر لك مواطن القوّة في أدائك.					
٢٥) قد حسّن إدارتك للانشطة داخل الفصل.					
٢٦) أثر على طريقة تعاملك مع مشاكل الطلبة السلوكية وانضباط الطلبة.					
٢٧) أثر على قدرتك على تحفيز (إشراك)الطلبة في العملية التعليمية					
٢٨) أثر على قدرتك في التعامل مع الفروق الفردية.					
٢٩) كان له تأثير في تقييمك الطلبة بشكل مستمر.					
٣٠) حسّن قدرتك على إعطاء الطلبة تغذية راجعة فعّالة.					
٣١) كان له تأثير في المكافأة الوظيفية مثل علاوة مالية أو مكافأة مالية سنوية.					
٣٢) ساهم في الترقية الوظيفية.					

ملاحظة هامة: اذا كان لديك الرغبة للمشاركة بحلقة نقاشية في المدرسة، تتضمن بعض المعلمين لعرض نظام تقويم معلم جديد، ومن خلال هذه الحلقة النقاشية تستطيع التعرّف على النظام الجديد المقترح، ويمكنك عرض وجهة نظرك بشكل أكبر تجاه هذا النظام.

الرجاء كتابة الاسم اوالبريد الالكتروني حتى يتمكن الباحث من تنظيم ذلك .....

او ارسال الاسم واسم المدرسة الى الرقم ٠٠٩٦٥٥٠٦٦٠٠٩٤ او من خلال البريد الالكتروني Talal.Almutairi@durham.ac.uk

Appendix (4): Information Sheet for Interview and Focus Group [English]

Information Sheet For This Research

The aims of the study:

The aim of this study is firstly to analyse and evaluate the current teacher evaluation system in Kuwait using data from teachers, head teachers and inspectors. The objectives are:

e) To determine what purposes dominate the current system and to compare this to participants' desired purposes for teacher evaluation.
f) To identify the tools that are used in the current system, and to compare this to tools of teacher evaluation that participants think *should* be used
g) To analyse the role of evaluators and how this is regarded by teachers in the current system.
h) To find out if and how the current teacher evaluation system is supporting the development of teachers' performance.

The second aim of the study is to suggest an alternative teachers evaluation system based on 'Risk-Based Analysis' approach to participants, and in order to probe their opinions about its potential for development and improvement of teacher evaluation in Kuwait.

In order to explore the aim of this study, the researcher relies on the data collected through focus groups and interviews.

a) The first interview will take approximately an hour and will be held with the head teacher and inspector to discuss the current teacher evaluation system. The second interview pertains to the alternative system and will take approximately an hour. Audio recording will be used to collect the data, unless the participant does not give her/his consent, in which case the researcher will take notes.

b) The focus groups will be used to present the alternative system to teachers in order to achieve the research aim. The focus groups will take place over two days and will last approximately two hours in total – one hour on each day. On the first day, the researcher will present the alternative system to the participants by providing them with a booklet. On the second day, the researcher will collect the teachers' views on the alternative system. Audio recording will be used to collect the data, unless the participants do not give their consent, in which case the researcher will take notes.

If you decide to participate in this study, you will be asked to sign a consent form. All information gathered during this research project will be treated confidentially and individual names will not be used at any point, thus guaranteeing your anonymity.

Researcher Details:
Name of the researcher:
Talal S. Almutairi
Position and contact information:
PhD Student
School of Education
Durham University
Durham, England
DH1 1 TA
UK
E-mail: Talal.Almutairi@durham.ac.uk
Mobile: +44 (0)7450020014 / +965 50660094

Appendix (5): Information Sheet for Interview and Focus Group [Arabic]

معلومات توضحية للدراسة

الهدف من الدراسة:

الهدف الرئيسي للدراسة هو تحليل وتقويم " نظام تقويم المعلم" في الكويت من خلال وجهات النظر المختلفة لأصحاب الميدان التربوي من مدراء وتوجيه فني ومعلمين:

ا) لتحديد اغراض التقويم في الواقع العملي والتطرق الى الاغراض ذات الاهمية الواجب استخدامها للنظام تقويم المعلم.
ب) لتحديد الادوات المستخدمة في نظام التقويم الحالي وماهي الادوات التي يجب استخدامها بنظام تقويم المعلم.
ج) التحقق من دور كل مقوم على حده في علميات تقويم المعلم.
د) التحقق من نظام تقويم المعلم لمعرفة ما اذا كان داعم لمعلم في تطوير وتحسيين ادائه.

وايضا الهدف من الدراسة عرض نظام بديل لتقويم المعلم على أصحاب الميدان التربوي للحصول على أرائهم حول نظام بديل يعتمد على "التحليل القائم على الخطورة" لمعرفة ما إذا كان هذا النظام يدفع الى تطوير نظام تقويم المعلم في دولة الكويت. والغرض هنا لتحديد الصعوبات حول النظام المقترح واكتشاف تباين الأراء بين اصحاب الميدان للتغلب بالنهاية على المشاكل والوصول الى مايرضي احتياجات اهل الميدان.

لتحقيق الاهداف المرجوه، الباحث يعتمد على جمع المعلومات من خلال عدة ادوات (حلقات المناقشة الجماعية والمقابلات)

ا) المقابلات الشخصية مع مدراء المدارس والموجهين. المقابلات الشخصية تنقسم الى قسمين: القسم الاول عن النظام الحالي والقسم الثاني عن النظام البديل وكل مقابلة تستغرق قرابة الساعة.

ب) الحلقات النقاشية مع المعلمين على فترات مختلفة لجمع المعلومات عن النظام البديل. الحلقات النقاشية متوقع ان تستغرق قرابة الساعتين. بالفترة الاولى سوف يتم تقديم مذكرة توضحيه للنظام البديل مع الشرح والنقاش التوضيحي. بالفترة الثانية سوف يتم جمع وجهات نظر المعلمين حول النظام البديل.

اذا قررت المشاركة في هذه الدراسة، يرجى تزويدنا بالموافقة وذلك من خلال التوقيع على مسودة المشاركة بالبحث. ونود ان نحيطكم علما ان جميع المعلومات الشخصية سوف تكون سرية والاسماء الحقيقية للمشاركين لن تستخدم ضمان للسرية.

اسم الباحث: طلال سعد المطيري
طالب دكتوراه في كلية التربية
جامعة درم / المملكة المتحدة
الايميل: Talal.Almutairi@durham.ac.uk
التليفون: ٠٠٩٦٥٥٠٦٦٠٠٩٤/٠٠٤٤٧٤٥٠٠٢٠٠١٤

Appendix (6): Consent Form for Interview and Focus Group [English]

Consent Request Form

Teacher Evaluation in the State of Kuwait

(The participant should complete the whole of this sheet himself/herself)

1) Have you read the Participant Information Sheet? Yes / No

2) Have you had an opportunity to ask questions and to discuss the study? Yes / No

3) Have you received satisfactory answers to all of your questions? Yes / No

4) Have you received enough information about the study? Yes / No

5) Do you consent to participate in the study? Yes / No

6) Do you understand that you are free to withdraw from the study: Yes / No

*At any time and without having to give a reason for withdrawing and without affecting your position.

Note:

I understand that my participation will be through **(Interview OR Focus Group)**

The interviews/Focus group will be recorded (agree / disagree)

I understand that all the views and information I provide will remain anonymous and will be treated confidentially. I understand that any information I provide will be stored securely.

Signed ..... Date .....

(NAME) .....

Appendix (7): Consent Form for Interview and Focus Group [Arabic]

نموذج الموافقة على المشاركة في البحث

١) هل قرأت المسودة الخاصة بالمشاركين التي تحتوي على معلومات البحث ؟ نعم / لا
٢) هل حصلت على الفرصة للسؤال عن موضوع البحث او مناقشة بعض النقاط ؟ نعم / لا
٣) هل حصلت على الاجوبة الكافية حول اسألتك ؟ نعم / لا
٤) هل حصلت عن المعلومات الكافية حول موضوع الدراسة ؟ نعم / لا
٥) هل توافق على المشاركة بالبحث ؟ نعم / لا
٦) هل تم اخبارك انه لك الحق بالانسحاب بأي وقت وبدون ذكر السبب ؟ نعم / لا

- هل توافق على تسجيل المقابلة صوتياً حتى يتمكن الباحث من تدوين ذلك لاحقاً مع التعهد الكامل بسرية المعلومات واستخدامها لأغراض البحث فقط . مع العلم أن الباحث لن يذكر الأسم أو أي معلومات شخصية، وأن التسجيل سوف يتم حذفه بعد تفريغ المعلومات.

الاسم ..... التاريخ .....
التوقيع .....

Appendix (8): Letter from the School of Education at Durham University to the Embassy of Kuwait in London Regarding this Research.

Durham University
School of Education

Shaped by the past, creating the future

Ref: 000229651

14th February 2014

TO WHOM IT MAY CONCERN

Re: Talal Almutairi (27/10/1985)

Talal Almutairi is a current Doctoral student in the School of Education at Durham University.

Mr Almutairi's research is about Teacher Evaluation in the state of Kuwait. He intends to start data collection from Public Schools in the Kuwaiti Educational Districts. Therefore, he would like to obtain permission for accessing schools to conduct the research with Teachers, Head Teachers, and Inspectors in the Ministry of Education in Kuwait.

Please do not hesitate to contact me at p.m.kind@durham.ac.uk or 0191 3348310 if you have any queries relating to the above.

Yours sincerely

PMKind

Dr Per Kind
Divisional Director for Postgraduate Research

UNIVERSITY of DURHAM · SCHOOL of EDUCATION

Durham University Durham City DH1 1TA
Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk/education
Durham University is the trading name of the University of Durham

Appendix (9): Letter from The School of Education at Durham University Shows Period for Conducting this Study.

Durham University
School of Education

Shaped by the past, creating the future

Ref: 000229651

17th February 2014

TO WHOM IT MAY CONCERN

Re: Talal Almutairi (27/10/1985)

Talal Almutairi is a current Doctoral student in the School of Education at Durham University. Mr Almutairi's research is about Teacher Evaluation in the state of Kuwait. He will undertake departmentally approved field work in public schools in the Kuwaiti Educational Districts for his PhD programme from 19 February 2014 to 10 June 2014.

Please do not hesitate to contact me at p.m.kind@durham.ac.uk or 0191 3348310 if you have any queries relating to the above.

Yours sincerely

Dr Per Kind
Divisional Director for Postgraduate Research

UNIVERSITY of DURHAM · SCHOOL of EDUCATION

Durham University Durham City DH1 1TA
Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk/education
Durham University is the trading name of the University of Durham

Appendix (10): Letter from the Embassy of Kuwait in London to the Ministry of Education in Kuwait for Granting Permission to Conduct this Study [Arabic]

Embassy of the State of Kuwait
Cultural Office
London

سـفارة دولـة الكـويت
المكتب الثقافي
لنـدن

م ق/G531
19 فبراير 2014م

إلى من يهمه الأمر

الموضوع: الرحلة العلمية للسيد/طلال سعد شريد المطيري
(ر.م.: 285102701459)

نفيدكم بان المذكور أعلاه والذي أبتعث من قبل ديوان الخدمة المدنية، للحصول على درجة الدكتوراه في تخصص " دراسات علم التربية: التقويم والقياس" بجامعة 'Duhram' إعتباراً من 2012/10/01م.

سيقوم برحلة علمية إلى دولة الكويت لجمع البيانات الإحصائية الخاصة ببحث الدكتوراه، وهذا يتطلب تطبيق استمارة البحث في مدارس وزارة التربية في جميع المناطق التعليمية في دولة الكويت.

كما نفيدكم أيضاً بأن الدراسة التي سيجريها المذكور هي عبارة عن تقييم لنظام تقويم المعلم بدولة الكويت. وهذا البحث هو عبارة عن تطبيق إستبيانات من أجل الحصول على المعلومات الرئيسية للدراسة، وايضاً إجراء مقابلات شخصية مع المعلمين ورؤساء الأقسام ومدراء المدارس والموجهين الفنيين.

نرجو من سيادتكم السماح للمذكور بإجراء وتطبيق هذه الدراسة وتقديم العون اللازم له وتسهيل مهمته البحثية حتى يستطيع الحصول على المعلومات والتي يمكن أن تحقق الهدف الرئيسي للدراسة . مرفق خطاب الجامعة والذي يؤكد أهمية إجراء هذه الدراسة والتي هي الأساس للحصول على درجة الدكتوراه.

نشكركم على حسن تعاونكم،،
وتقبّلوا فائق التقدير والاحترام،،،

أ.د محمد حمد الحمد الهاجري
رئيس المكتب الثقافي

محاسن القدال

60A Knightsbridge, London SW1X 7JX
Please quote our reference
Telephone: 020-7761 8500 Fax: 020-7761 8505
www.kuwaitculturaloffice.org.uk

Appendix (11): Letter from the Director of the Educational Research Department to the Asimah Educational District [Arabic]

75 / 30 / EA0098394/05/5000

الرقم: ٥٠٤

التاريخ: ١٧/٢/٢٠١٤م

السيد المحترم / مدير عام منطقة العاصمة التعليمية

تحية طيبة وبعد ،،،

الموضوع / تسهيل مهمة

يقوم الطالب / طلال سعد شريد المطيري المسجل في جامعة درهام بالمملكة المتحدة (لندن) بإجراء دراسة بعنوان " تقييم لنظام تقويم المعلم بدولة الكويت ".

فيرجى تسهيل مهمة المذكور أعلاه لعمل البحث الميداني من خلال عمل مقابلات شخصية وتطبيق اداة البحث استبانه المختومة صفحاتها من إدارة البحوث التربوية على عينه من مدراء مدارس و معلمين وموجهين فنيين وطلاب مدارس المرحلة الابتدائية بنين - بنات التابعة لمنطقتكم التعليمية خلال الفصل الدراسي الحالي 2014/2013.

مع خالص الشكر والتقدير

مدير إدارة البحوث التربوية

أ. إبتسام الحمادي
مدير إدارة البحوث التربوية

وزارة التربية
إدارة البحوث التربوية

Aisha -
- نسخة للملف

ص.ب : ١٦٢٢٢ القادسية - ٣٥٨٥٣ الكويت - تلفون : ٤٨٣٨٣٢١ - ٤٨٤٢٤٠٤ - فاكس : ٤٨٣٧٩٠٩ - ٤٨٤٢٤٠٤
P.O. Box : 16222 - QADSIAH - 35853 - KUWAIT - Tel. : 4842404 - 4838321 - Fax : 4837909 - 4842404

Appendix (12): Letter from the Director of the Educational Research Department to the Ahmadi Educational District [Arabic]

75 / 30 / EA0098394/05/5000

الرقم: ٥٥٢

التاريخ: ٢٠١٤/٢/١٧

السيد المحترم/ مدير عام منطقة الأحمدي التعليمية

تحية طيبة وبعد ،،،

الموضوع/ تسهيل مهمة

يقوم الطالب / طلال سعد شريد المطيري المسجل في جامعة درهام بالمملكة المتحدة (لندن) باجراء دراسة بعنوان "تقييم لنظام تقويم المعلم بدولة الكويت".

فيرجى تسهيل مهمة المذكور أعلاه لعمل البحث الميداني من خلال عمل مقابلات شخصية وتطبيق اداة البحث استبانه المختومة صفحاتها من إدارة البحوث التربوية على عينه من مدراء مدارس و معلمين وموجهين فنيين وطلاب مدارس المرحلة الابتدائية بنين - بنات التابعة لمنطقتكم التعليمية خلال الفصل الدراسي الحالي 2014/2013.

مع خالص الشكر والتقدير

مدير إدارة البحوث التربوية

أ. إبتسام العماي
مدير إدارة البحوث التربوية

وزارة التربية
إدارة البحوث التربوية

Aisha -
- نسخة للملف

ص.ب : ١٦٢٢٢ القادسية - ٣٥٨٥٣ الكويت - تلفون : ٤٨٣٨٣٢١ - ٤٨٤٢٤٠٤ - فاكس : ٤٨٣٧٩٠٩ - ٤٨٤٢٤٠٤
P.O. Box : 16222 - QADSIAH - 35853 - KUWAIT - Tel. : 4842404 - 4838321 - Fax : 4837909 - 4842404

Appendix (13): Letter from the Director of the Educational Research Department to the Farwaniya Educational District [Arabic]

75 / 30 / EA0098394/05/5000

الرقم: ٥٥٥ التاريخ: ٢٧/ ٢/ ٢٠١٤

السيد المحترم / مدير عام منطقة الفروانية التعليمية

تحية طيبة وبعد ،،،

الموضوع / تسهيل مهمة

يقوم الطالب / طلال سعد شريد المطيري المسجل في جامعة درهام بالمملكة المتحدة (لندن) بإجراء دراسة بعنوان "تقييم لنظام تقويم المعلم بدولة الكويت".

فيرجى تسهيل مهمة المذكور أعلاه لعمل البحث الميداني من خلال عمل مقابلات شخصية وتطبيق اداة البحث استبانه المختومة صفحاتها من إدارة البحوث التربوية على عينه من مدراء مدارس و معلمين وموجهين فنيين وطلاب مدارس المرحلة الابتدائية بنين - بنات التابعة لمنطقتكم التعليمية خلال الفصل الدراسي الحالي 2014/2013.

مع خالص الشكر والتقدير

مدير إدارة البحوث التربوية

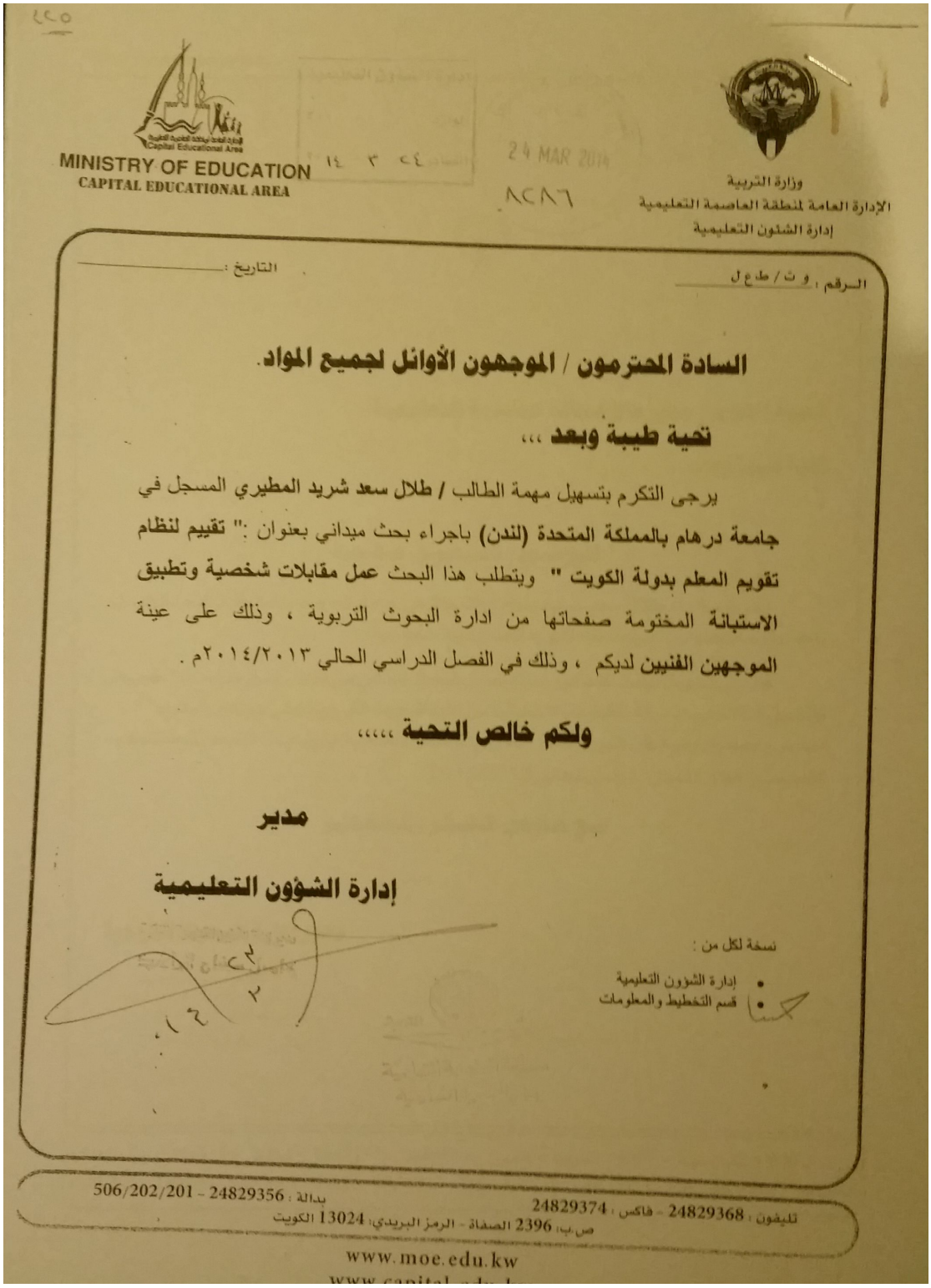
أ. ابتسام الحماي
مدير إدارة البحوث التربوية

وزارة التربية
إدارة البحوث التربوية

Aisha -
- نسخة للملف

ص.ب : ١٦٢٢٢ القادسية - ٣٥٨٥٣ الكويت - تلفون : ٤٨٣٨٣٢١ - ٤٨٤٢٤٠٤ - فاكس : ٤٨٣٧٩٠٩ - ٤٨٤٢٤٠٤
P.O. Box : 16222 - QADSIAH - 35853 - KUWAIT - Tel. : 4842404 - 4838321 - Fax : 4837909 - 4842404

Appendix (14): Letter from the Director of the Asimah Educational District to the Head Inspectors in the District [Arabic]



MINISTRY OF EDUCATION
CAPITAL EDUCATIONAL AREA

وزارة التربية
الإدارة العامة لمنطقة العاصمة التعليمية
إدارة الشئون التعليمية

24 MAR 2014

الرقم : و ت / ط ع ل

التاريخ :

السادة المحترمون / الموجهون الأوائل لجميع المواد.

تحية طيبة وبعد ،،،

يرجى التكرم بتسهيل مهمة الطالب / **طلال سعد شريد المطيري** المسجل في **جامعة درهام بالمملكة المتحدة (لندن)** باجراء بحث ميداني بعنوان :" **تقييم لنظام تقويم المعلم بدولة الكويت** " ويتطلب هذا البحث **عمل مقابلات شخصية وتطبيق الاستبانة** المختومة صفحاتها من ادارة البحوث التربوية ، وذلك على عينة **الموجهين الفنيين** لديكم ، وذلك في الفصل الدراسي الحالي ٢٠١٣/٢٠١٤م .

ولكم خالص التحية ،،،،

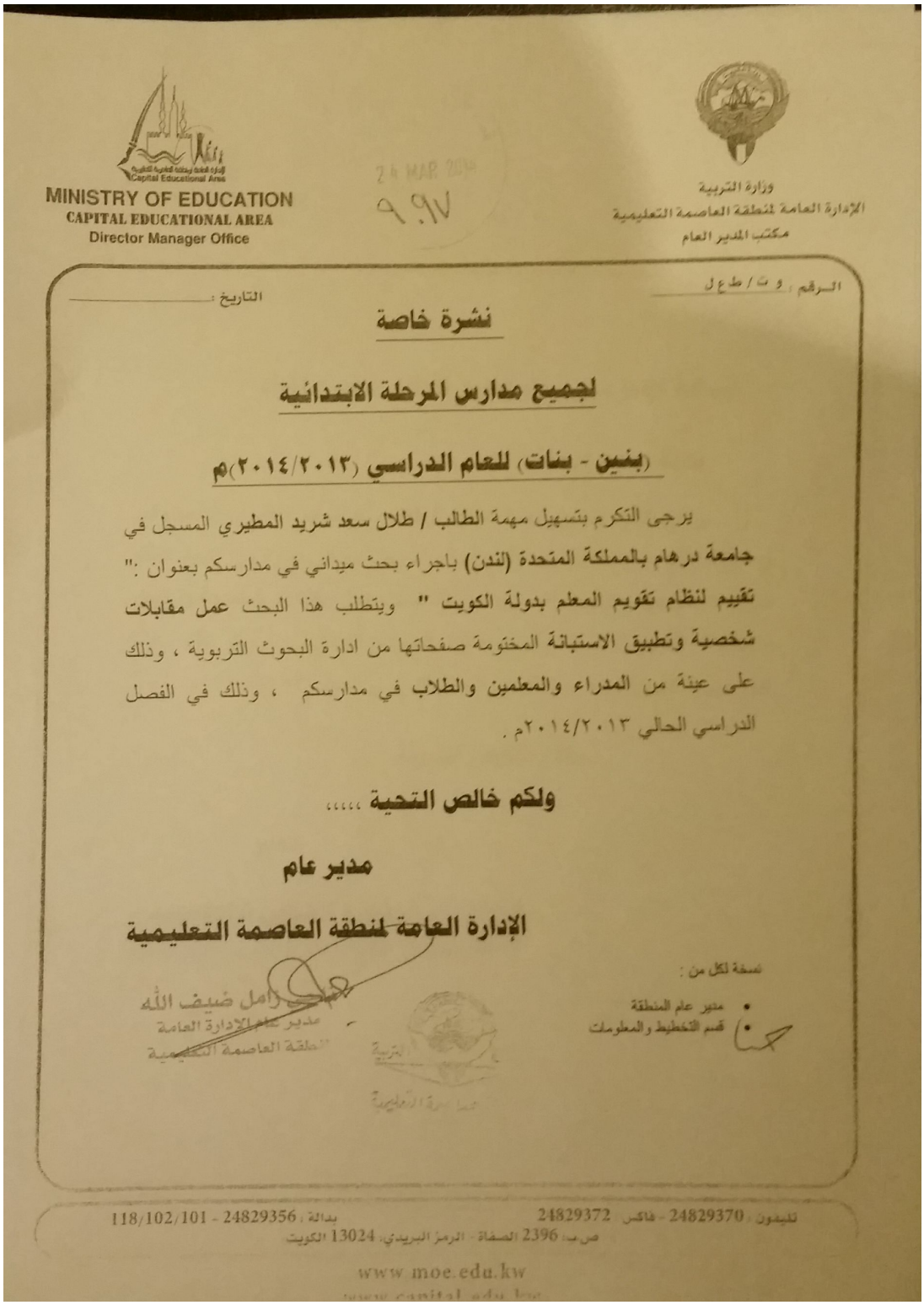
مدير

إدارة الشؤون التعليمية

نسخة لكل من :
- إدارة الشؤون التعليمية
- قسم التخطيط والمعلومات

برقية : 506/202/201 - 24829356
تليفون : 24829368 - فاكس : 24829374
ص.ب: 2396 الصفاة - الرمز البريدي: 13024 الكويت
www.moe.edu.kw

Appendix (15): Letter from the Director of the Asimah Educational District to the Head Teachers in the District [Arabic]



MINISTRY OF EDUCATION
CAPITAL EDUCATIONAL AREA
Director Manager Office

وزارة التربية
الإدارة العامة لمنطقة العاصمة التعليمية
مكتب المدير العام

24 MAR 2014
٩٩٧

الرقم : و ت / ط ع ل

التاريخ : _____

نشرة خاصة

لجميع مدارس المرحلة الابتدائية

(بنين - بنات) للعام الدراسي (٢٠١٣/٢٠١٤)م

يرجى التكرم بتسهيل مهمة الطالب / **طلال سعد شريد المطيري** المسجل في **جامعة درهام بالمملكة المتحدة (لندن)** باجراء بحث ميداني في مدارسكم بعنوان :" **تقييم لنظام تقويم المعلم بدولة الكويت** " ويتطلب هذا البحث **عمل مقابلات شخصية وتطبيق الاستبانة** المختومة صفحاتها من ادارة البحوث التربوية ، وذلك على عينة من **المدراء والمعلمين والطلاب** في مدارسكم ، وذلك في الفصل الدراسي الحالي ٢٠١٣/٢٠١٤م .

ولكم خالص التحية ،،،،

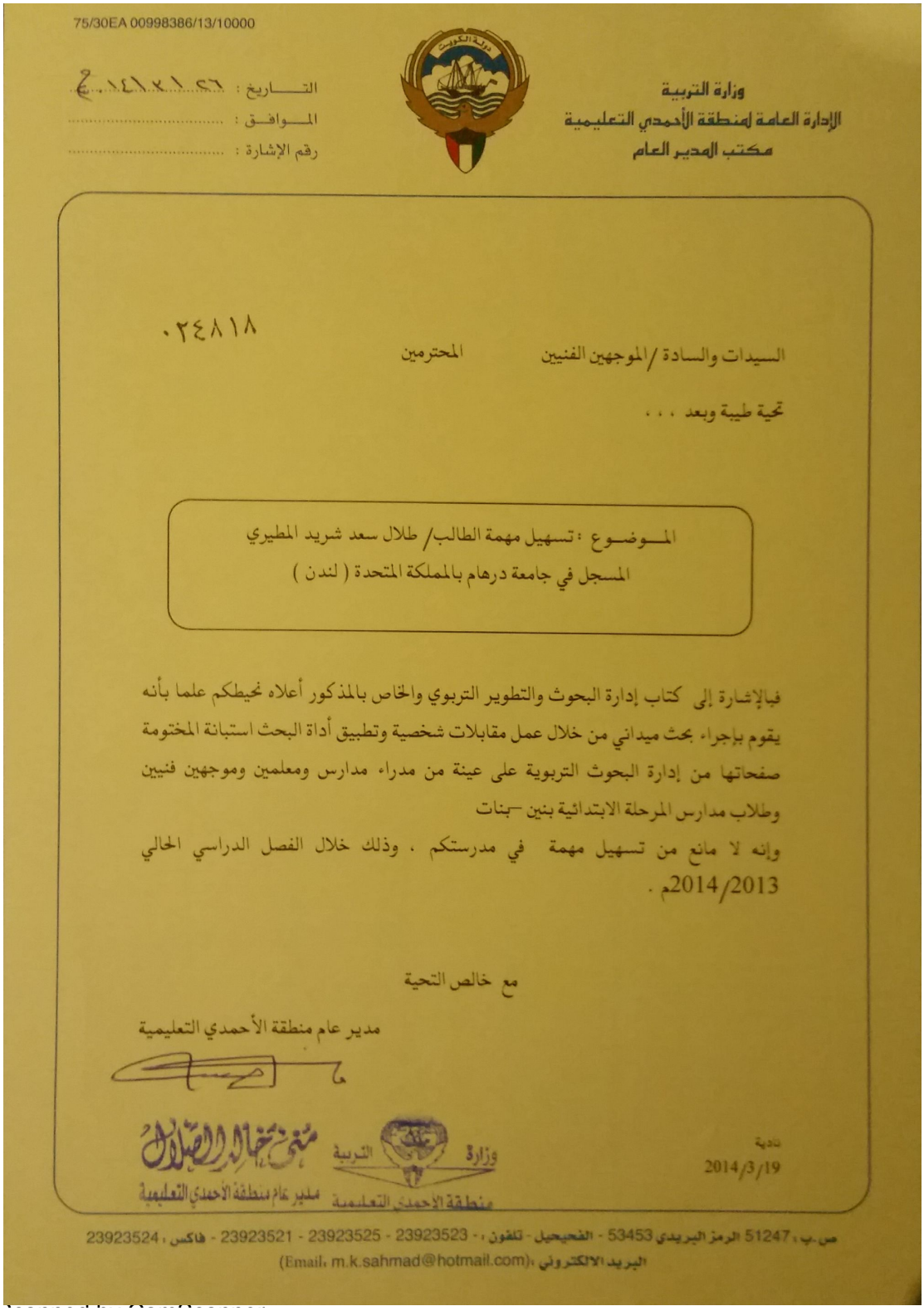
مدير عام

الإدارة العامة لمنطقة العاصمة التعليمية

نسخة لكل من :
- مدير عام المنطقة
- قسم التخطيط والمعلومات

تليفون : 24829370 - فاكس : 24829372 بدالة : 24829356 - 118/102/101
ص.ب: 2396 الصفاة - الرمز البريدي: 13024 الكويت
www.moe.edu.kw

Appendix (16): Letter from the Director of the Ahmadi Educational District to the Head Inspectors in the District [Arabic]



75/30EA 00998386/13/10000

التـــاريخ : ٢٠١٤/٨/٢٦
المـــوافـق : .....
رقم الإشارة : .....

وزارة التربية
الإدارة العامة لمنطقة الأحمدي التعليمية
مكتب المدير العام

٠٢٤٨١٨

السيدات والسادة /الموجهين الفنيين المحترمين

تحية طيبة وبعد . . .

المـــوضـــوع : تسهيل مهمة الطالب/ طلال سعد شريد المطيري
المسجل في جامعة درهام بالمملكة المتحدة (لندن)

فبالإشارة إلى كتاب إدارة البحوث والتطوير التربوي والخاص بالمذكور أعلاه نحيطكم علما بأنه يقوم بإجراء بحث ميداني من خلال عمل مقابلات شخصية وتطبيق أداة البحث استبانة المختومة صفحاتها من إدارة البحوث التربوية على عينة من مدراء مدارس ومعلمين وموجهين فنيين وطلاب مدارس المرحلة الابتدائية بنين –بنات

وإنه لا مانع من تسهيل مهمة في مدرستكم ، وذلك خلال الفصل الدراسي الحالي 2014/2013م .

مع خالص التحية

مدير عام منطقة الأحمدي التعليمية

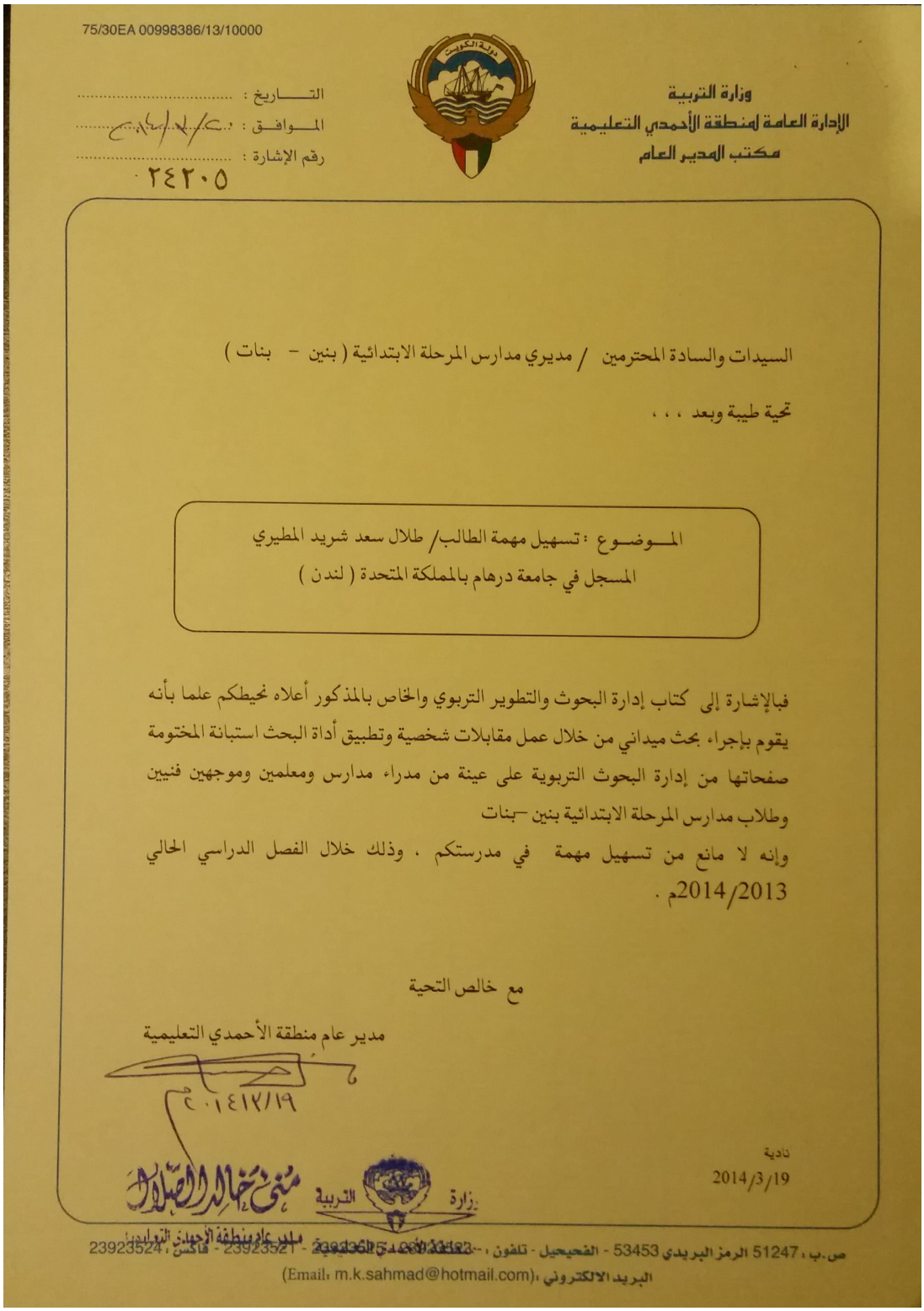
منى خالد الصلال
مدير عام منطقة الأحمدي التعليمية

وزارة التربية
منطقة الأحمدي التعليمية

بتاريخ
2014/3/19

ص.ب : 51247 الرمز البريدي 53453 - الفحيحيل - تلفون : - 23923523 - 23923525 - 23923521 - فاكس : 23923524
البريد الالكتروني : (Email: m.k.sahmad@hotmail.com)

Appendix (17): Letter from the Director of the Ahmadi Educational District to the Head Teachers in the District [Arabic]



75/30EA 00998386/13/10000

التـــــاريخ : .....
المـــوافـــق : .....
رقم الإشارة : .....
٢٤٢٠٥

وزارة التربية
الإدارة العامة لمنطقة الأحمدي التعليمية
مكتب المدير العام

السيدات والسادة المحترمين / مديري مدارس المرحلة الابتدائية (بنين - بنات)

تحية طيبة وبعد ، ، ،

المـــوضـــوع : تسهيل مهمة الطالب/ طلال سعد شريد المطيري
المسجل في جامعة درهام بالمملكة المتحدة (لندن)

فبالإشارة إلى كتاب إدارة البحوث والتطوير التربوي والخاص بالمذكور أعلاه نحيطكم علما بأنه يقوم بإجراء بحث ميداني من خلال عمل مقابلات شخصية وتطبيق أداة البحث استبانة المختومة صفحاتها من إدارة البحوث التربوية على عينة من مدراء مدارس ومعلمين وموجهين فنيين وطلاب مدارس المرحلة الابتدائية بنين –بنات

وإنه لا مانع من تسهيل مهمة في مدرستكم ، وذلك خلال الفصل الدراسي الحالي 2014/2013م .

مع خالص التحية

مدير عام منطقة الأحمدي التعليمية

١٤٣٥/٦/١٩ هـ

منى خالد الصالح

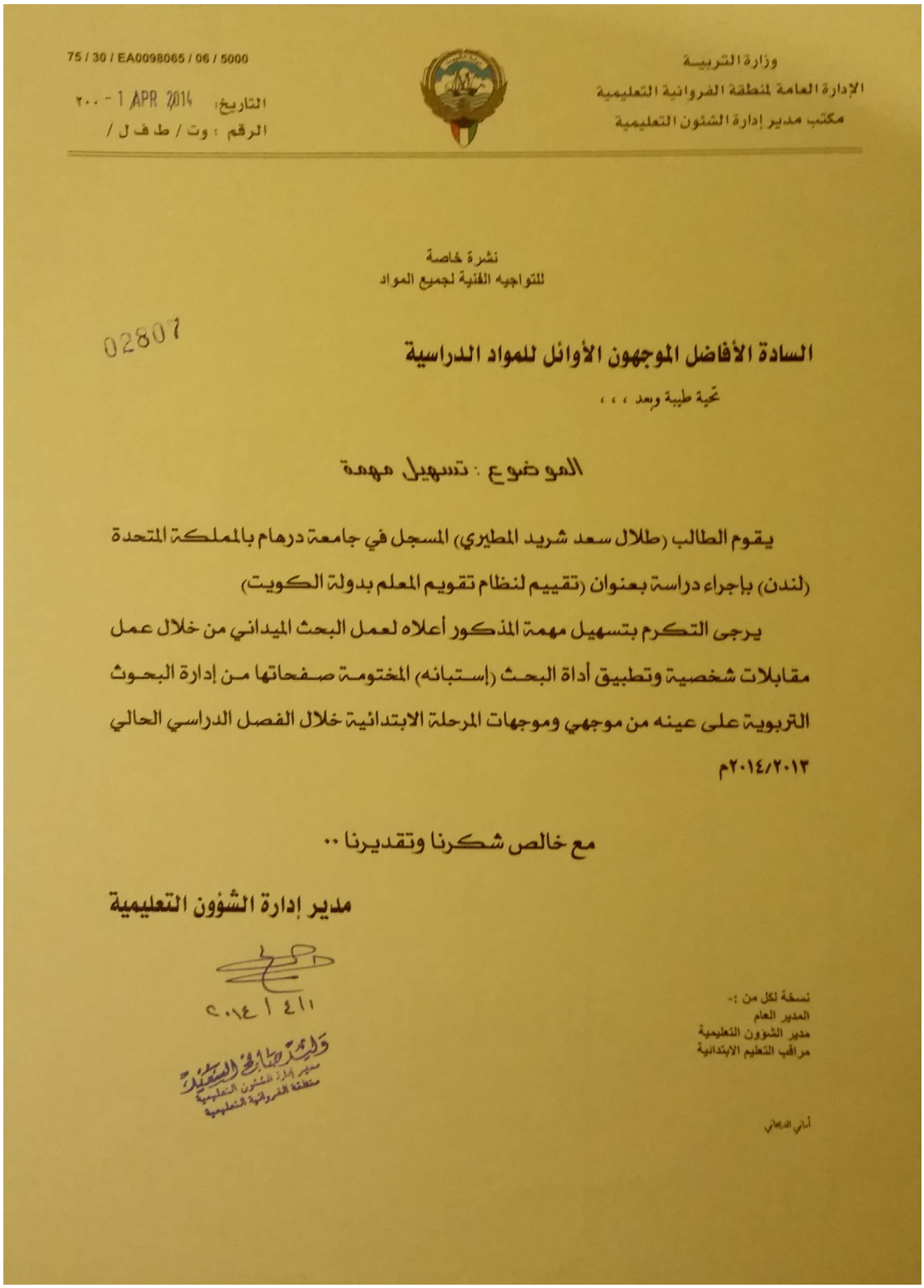
وزارة التربية

مدير عام منطقة الأحمدي التعليمية

نادية
2014/3/19

ص.ب ، 51247 الرمز البريدي 53453 - الفحيحيل - تلفون ، - 23923524 ، فاكس - 23923521
البريد الالكتروني ، (Email, m.k.sahmad@hotmail.com)

Appendix (18): Letter from the Director of the Farwaniya Educational District to the Head Inspectors in the District [Arabic]



75 / 30 / EA0098065 / 06 / 5000

التاريخ: ٢٠٠ - 1 APR 2014

الرقم : وت / ط ف ل /

وزارة التربيـــة
الإدارة العامة لمنطقة الفروانية التعليمية
مكتب مدير إدارة الشئون التعليمية

نشرة خاصة
للتواجيه الفنية لجميع المواد

02807

السادة الأفاضل الموجهون الأوائل للمواد الدراسية

تحية طيبة وبعد ،،،

الموضوع : تسهيل مهمة

يقوم الطالب (طلال سعد شريد المطيري) المسجل في جامعة درهام بالمملكة المتحدة (لندن) بإجراء دراسة بعنوان (تقييم لنظام تقويم المعلم بدولة الكويت)

يرجى التكرم بتسهيل مهمة المذكور أعلاه لعمل البحث الميداني من خلال عمل مقابلات شخصية وتطبيق أداة البحث (إستبانه) المختومة صفحاتها من إدارة البحوث التربوية على عينه من موجهي وموجهات المرحلة الابتدائية خلال الفصل الدراسي الحالي ٢٠١٣/٢٠١٤م

مع خالص شكرنا وتقديرنا ..

مدير إدارة الشؤون التعليمية

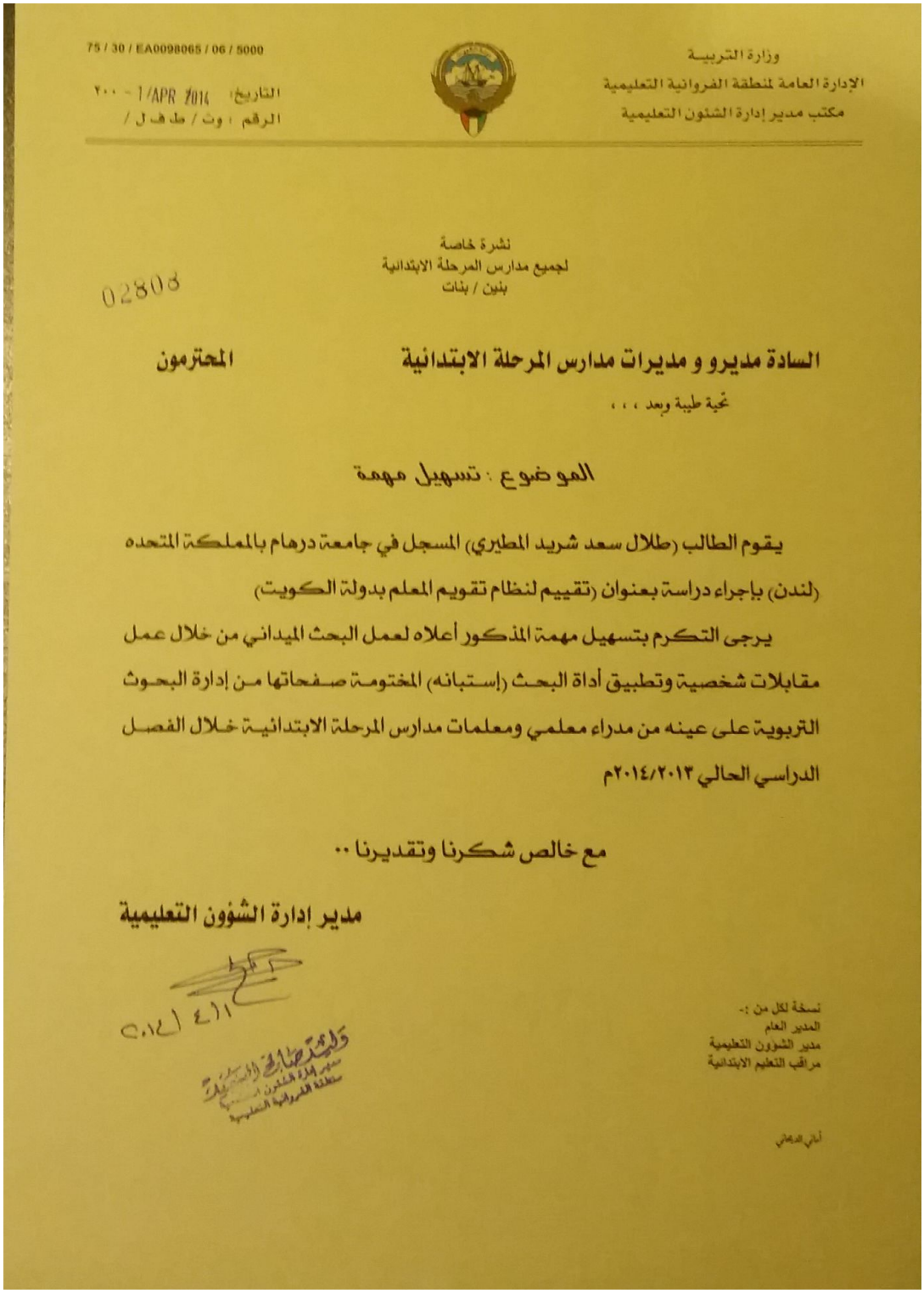
٢٠١٤ / ٤ / ١١

وليد طارق السعيد
مدير إدارة الشئون التعليمية
منطقة الفروانية التعليمية

نسخة لكل من :-
المدير العام
مدير الشؤون التعليمية
مراقب التعليم الابتدائية

أماني الدعاني

Appendix (19): Letter from the Director of the Farwaniya Educational District to the Head Teachers in the District [Arabic]



75 / 30 / EA0098065 / 06 / 5000

التاريخ: ٢٠١٤/APR/١٦ - ٢٠

الرقم : وت / ط ف ل /

وزارة التربيـــة
الإدارة العامة لمنطقة الفروانية التعليمية
مكتب مدير إدارة الشئون التعليمية

نشرة خاصة
لجميع مدارس المرحلة الابتدائية
بنين / بنات

02803

السادة مديرو و مديرات مدارس المرحلة الابتدائية المحترمون

تحية طيبة وبعد ،،،

الموضوع : تسهيل مهمة

يقوم الطالب (طلال سعد شريد المطيري) المسجل في جامعة درهام بالمملكة المتحده (لندن) بإجراء دراسة بعنوان (تقييم لنظام تقويم المعلم بدولة الكويت)

يرجى التكرم بتسهيل مهمة المذكور أعلاه لعمل البحث الميداني من خلال عمل مقابلات شخصية وتطبيق أداة البحث (إستبانه) المختومة صفحاتها من إدارة البحوث التربوية على عينه من مدراء معلمي ومعلمات مدارس المرحلة الابتدائية خلال الفصل الدراسي الحالي ٢٠١٣/٢٠١٤م

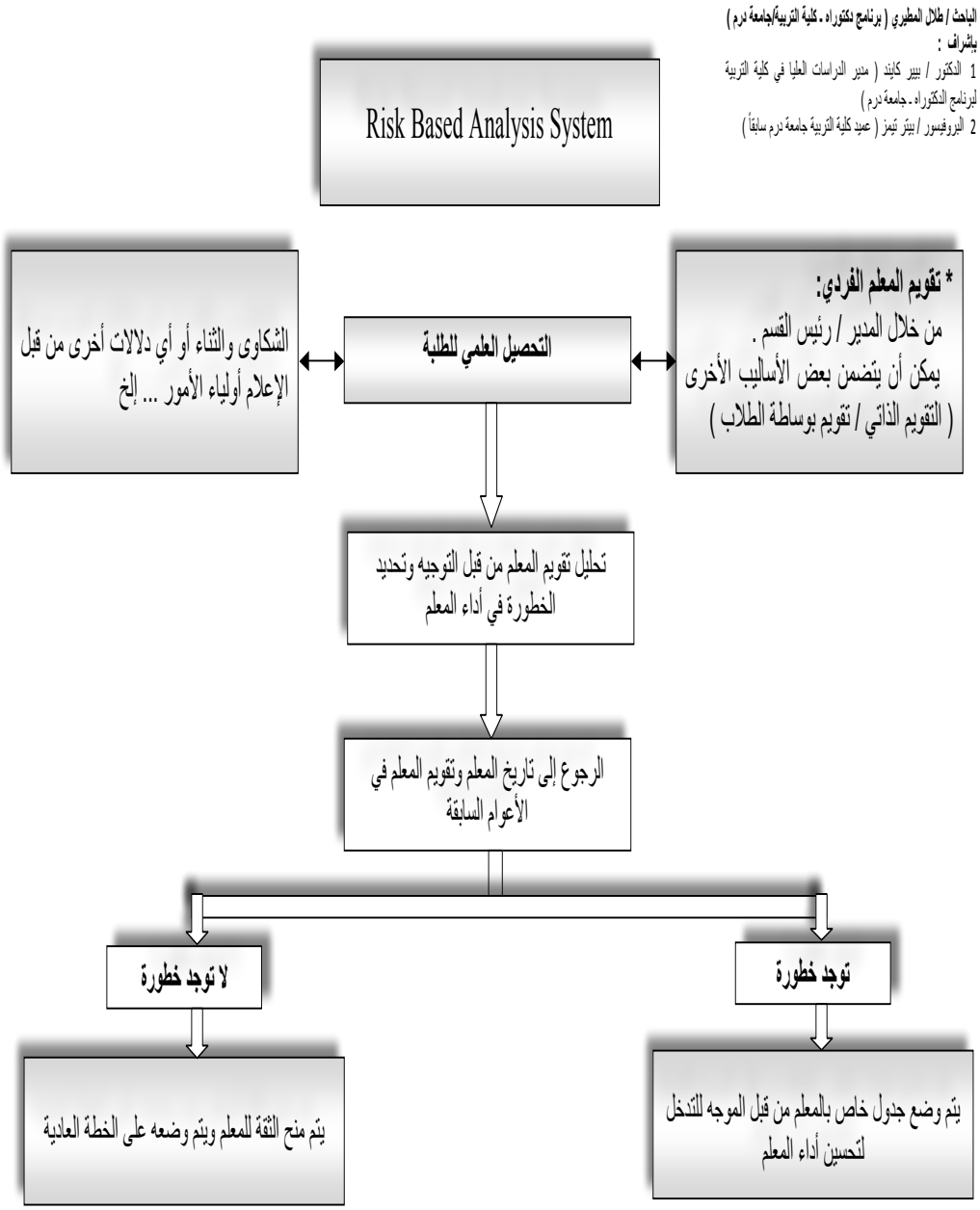
مع خالص شكرنا وتقديرنا ..

مدير إدارة الشؤون التعليمية

نسخة لكل من :-
المدير العام
مدير الشؤون التعليمية
مراقب التعليم الابتدائية

أماني الدعماني

Appendix (20): Alternative System Booklet that was Presented to the Participants [Arabic]



Inspectorate of Education, Netherlands (2008)

النظام البديل المقترح لتقويم المعلم

الخطوة الاولى: تقويم الاداء الفردي للمعلم
هذه الخطوة تبدأ في بداية كل عام الدراسي في شهر سبتمبر الى شهر مايو. جميع المعلمين يخضعون لتقويم الاداء قبل مرحلة تحديد الخطورة.

آ) تقويم اداء المعلم:

يتم تقويم كل معلم من قبل رئيس القسم ومدير المدرسة. ويكون الغرض من تقويم المعلم تعزيز التطوير المهني وتحسين الاداء. ايضا التقويم هنا يستخدم للاغراض التالية تحديد اداء المعلم خلال العام الدراسي واعطاء تقدير نهائي (مثلا: جيد ، جيد جدا ، ... الخ) واصدار احكام على اداء المعلم في العام الدراسي بشكل كامل مثل الكادر والترقية والفصل من العمل الخ.

يعتمد تقويم المعلم (في هذه المرحلة) على ادوات مختلفة لقياس اداء المعلم، كالتالي:

الملاحظة الصفية: يتم ملاحظة المعلم في الفصل من قبل المقوم. المقوم هنا لابد ان يستخدم صحفية الملاحظة الموحدة ولذلك لاعطاء المعلم التغذية الراجعة بالاضافة الى ارفاقها بالتقويم النهائى.

التقويم الذاتي: ان يسمح للمعلم بتقويم نفسه من خلال استخدام صحيفة رسمية موحدة.

تقويم القرناء او الزملاء: هنا يتم استخدام هذه الاداة للاغراض التكوينية بهدف التطوير المهني وليس لاتخاذ حكم على المعلم. ويكون التقويم اما من خلال الملاحظة الصفية او من خلال كتاب اعداد الدروس والمستندات المتعلقة بالاداء لاعطاء التغذية الراجعة.

الملف الانجازي للمعلم: يتم استخدام الملف الانجازي لاعمال المعلم خلال العام الدراسي ولذلك لمساعدة المقوم على اتخاذ حكم تجاة اعمال المعلم التي لايمكن ملاحظتها صفيا مثل تقويم المعلم للطلاب وتحسن الطلاب ومتابعتهم والانشطة المدرسية.

تقويم الطلبة: والمقصود هنا تقويم الطالب للاداء المعلم وهنا الحرية للمقوم باستخدام المقابلات او الاستبيان. وبالنسبة للمعلم ذو الفصول التعليمية الكثيرة، خمس فصول واكثر سوف يقوم المقوم باختيار بعض الفصول وليس جميع الفصول.

بالنسبة للمعايير المستخدمة لتقويم المعلم سوف تكون مماثلة للمعايير المستخدمة حاليا في نظام تقويم المعلم في دولة الكويت.ويكون ايضا الحكم وفق الدرجات التالية: ممتاز، جيد حدا ، جيد ، ضعيف مماثل لما هو مستخدم حاليا.

من خلال هذه المرحلة يحصل المعلم على تقرير الاداء في منتصف العام الدراسي وايضا تغذية راجعة بعد كل ملاحظة صفية. بعد شهر مايو من العام الدراسي، المعلم سوف يحصل على النتيجة النهائية (التقرير النهائي) بالتفصيل عن اداء المعلم الفردي بعد اعتماد النتيجة من قبل المقومين. المعلم سوف يحصل على النتيجة النهائية من خلال المنطقة التعليمية او ارسال مغلف النتيجة الى عنوان المعلم "السكن". التقرير سوف تتضمن: نصائح لتطوير الاداء، نقاط الضعف والقوى، جميع صحائف الملاحظة الصفية، والدرجة النهائية وفق المعايير.

ب) اختبارات قياس اداء الطلاب:
في شهر مايو، سوف يتم ارفاق نتائج اختبارات الموحدة لقياس الاداء للطلاب مع تقويم المعلم النهائي. الاختبارات سوف تطبق لجميع المواد التعليمية ذات المنهج الموحد على مستوى الدولة. المدارس هي المسؤولة عن تطبيق وتصميم واصدار النتائج ، سوف تصنف درجات الطلاب وفق المرحلة والمدرسة والصف ويتم فرزها لكل معلم على حده من قبل التوجيه بالمرحلة الثانية.

ج) الاشارات / المستندات

لابد ان يرفق ايضا مع التقويم النهائي للمعلم ملف خاص بالشكاوى والثناء من قبل اولياء الامور ان وجد. بالاضافة الى شهادات الدورات التدريبية وورش العمل، شهادات اخرى حصل عليها المعلم تصب في الجانب التعليمي من قبل بعض المؤسسات التعليمية او الجمعيات الثقافية او المهنية مثل جمعية المعلمين لاخذ بعين الاعتبار التطويرالمهني و الذاتي للمعلم خلال العام الدراسي.

المرحلة الثانية: الكشف لتحديد الخطورة

هذه المرحلة تبدآ في نهاية كل عام الدراسي في مايو او يونيو ، اقسام التوجيه في المناطق التعليمية سوف تصل لهم (أ) التقارير النهائية للتقويم المعلم ، بالاضافة الى نتائج (ب) اختبارات قياس اداء التلاميذ و(ج) المستندات/ الاشارات المتعلقة بالمعلم.

سوف يشكل فريق من الموجهين لكل مادة على حده لبدآ بعملية الكشف عن المخاطر استنادا على الادلة اعلاه. اذا تم الكشف على خطورة في اداء المعلم من قبل التوجيه مثال اذا وجد هناك عدم توافق بين درجة المعلم النهائية مع نتائج الاختبارات او هناك مشكلة تم اكتشافها من خلال المستندات المرفقة او اذا كانت النتيجة لاداء المعلم ضعيف او جيد. سوف يتم تحديد خطورة في اداء المعلم ، لذلك سوف يتم الرجوع الى تاريخ المعلم واذا وجد في تاريخ المعلم ان المعلم ممتاز في السنوات مثلا الخمس او العشر الماضية سوف يتم اعطاء المعلم فرصة لتحسين ادائه مع المقوم الداخلي للعام الدراسي المقبل ، اما اذا تاريخ المعلم لم يكن كذلك مثل حصل على جيد او ضعيف سابقا سوف يتم وضع المعلم على خطة التدخل من قبل التوجيه بالعام الدراسي المقبل من خلال موجه فني واحد بالاشتراك مع مدير المدرسة فقط.

واذا لم يتم تحديد خطورة سوف يتم وضع المعلم على الخطة العادية اي بمعنى المعلم سوف يتم تقويمه من قبل التوجيه بالاشتراك مع مدير المدرسة ورئيس القسم كل اربع او ثلاث سنوات، مالم يتم تحديد خطورة خلال هذه الفترة (مرحلة الكشف لتحديد الخطورة) في العام المقبل. وخلال هذه السنوات سوف يتم تقويم المعلم من قبل المقوم الداخل حتى يحين موعد الخطة العادية.

المرحلة الثالثة: التدخل من قبل التوجيه الفني

هذه المرحلة تبدا في العام الدراسي الذي يلي مرحلة الكشف اي بمعنى شهر سبتمبر من العام الدراسي. الموجه سوف يشارك في تقوم المعلم المشخص ضمن الخطورة حيث يقوم الموجه في تقويم المعلم في الجوانب التدريسية والتعليمية بينما مدير المدرسة يركز على تقويم المعلم بالالتزام بالدوام المدرسي والتعاون مع الادارة والطاقم التعليمي. وتكون مشاركته بدلا عن رئيس القسم في المرحلة الاولى (أ) تقويم اداء المعلم (ب) اختبار التلاميذ (ج) الاشارات. اي بمعنى اخر، الموجه سوف يشارك في تقويم المعلم بالادوات المختلفة المذكورة ويساعد في ارفاق اختبارات التلاميذ والاشارات حتي يتم ارسالها للتوجيه في مرحلة تحديد الخطورة القادمة.

خلال مرحلة التدخل سوف يكون دور الموجه شامل ايضا التحضير لبرنامج للتطوير وخطة علاجية مرافقة لتقويم المعلم على سبيل المثال وضع الدورات التطويرية وورش العمل المناسبة للمعلم ومتابعة اللتزام المعلم بذلك استنادا على تقويم المعلم بالعام الماضي الذي تم تحديد الخطورة من خلاله.

واذا لم يتطور المعلم من خلال التدخل من قبل موجه واحد وتم تحديد خطورة ايضا في مرحلة الكشف التالية لتدخل التوجية سوف يوضع المعلم على الخطة العلاجية المكثفة قبل اتخاذ قرار الفصل او التحويل من مهنة التدريس الى مهنة ادارية.خلال المرحلة التدخل المكثفة سوف يتم تقويم المعلم ومتابعته من قبل اثنين من التوجيه الفني للمادة التعليمية، بينما دور مدير المدرسة يقتصر على اعطاء تقرير عن التزام المعلم بالدوام المدرسي والتعاون مع المدرسة و الزملاء. واذا لم يتطور المعلم بعد ذلك سوف يتم اتخاذ العقوبات تجاه المعلم. وتكون ايضا هنا مشاركة المقومين من خلال استخدام ادوات مختلفة وارفاق التقارير لاختبارات الطلاب والاشارات للاستفادة منهم في مرحلة تحديد الخطورة التي تلي مرحلة التدخل المكثف في العام المقبل، لاثبات ان المعلم تطور او يستلزم ان يتخذ ضدة قرار التحويل او الفصل حسب قوانين الوزارة

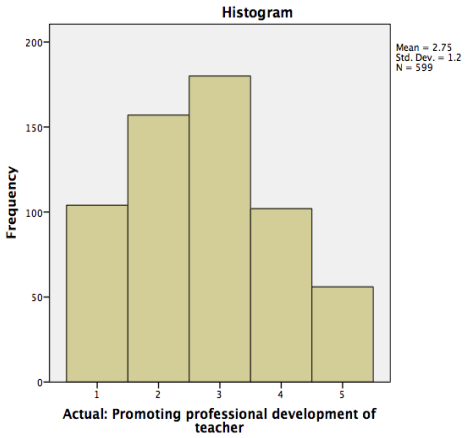
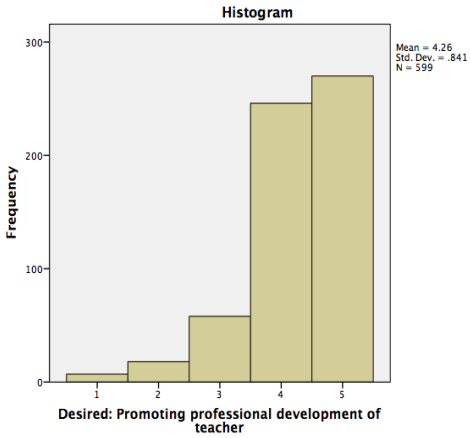
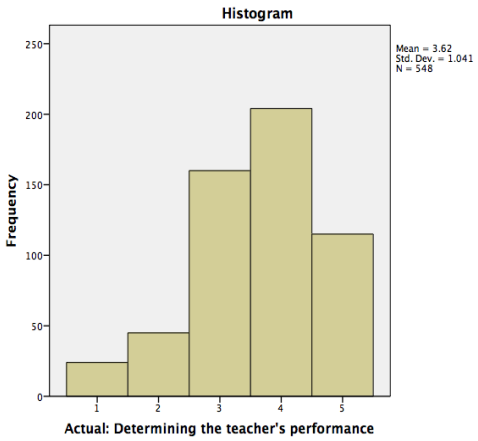
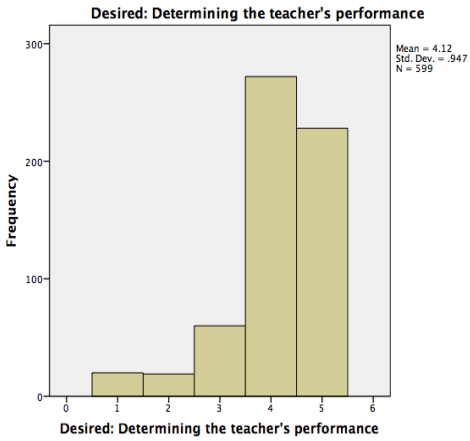
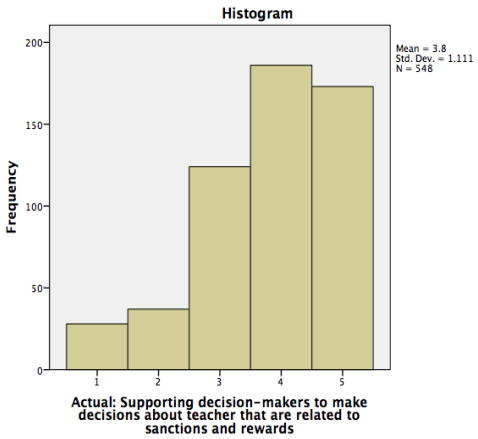
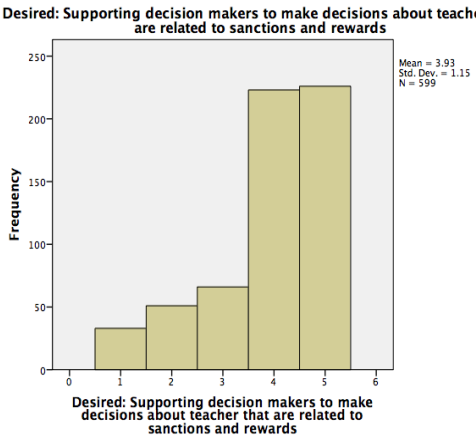
اما في مايتعلق بالتدخل للجميع المعلمين على الخطة العادية لتقويم جميع المعلمين بغض النظر عن الخطورة، سوف تكون من خلال الموجه ومدير المدرسة ورئيس القسم. وهنا جميع المقومين سوف يقومون بالاشتراك في تقويم المعلم

سواء فيما يتعلق بالجانب التدريسي او الجانب الاداري والتعاون في بيئة العمل. وهذه الخطة سوف تطبق كل اربعة اعوام او ثلاثة اعوام وفقا للخطوة الاولى، اي بمعنى ان التقويم يتضمن ادوات مختلفة ونتائج الاختبارات قياس للطلبة وارفاق المستندات والاشارات لمرحلة تحديد الخطورة.

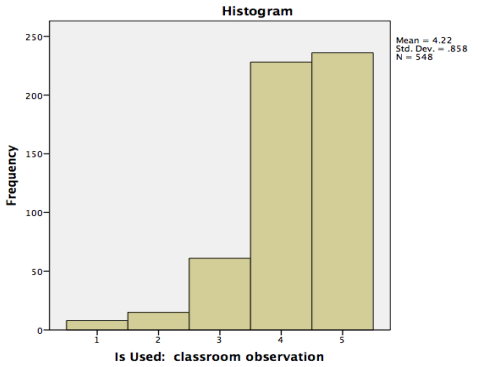
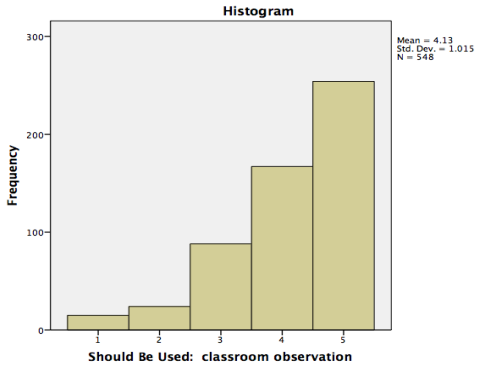
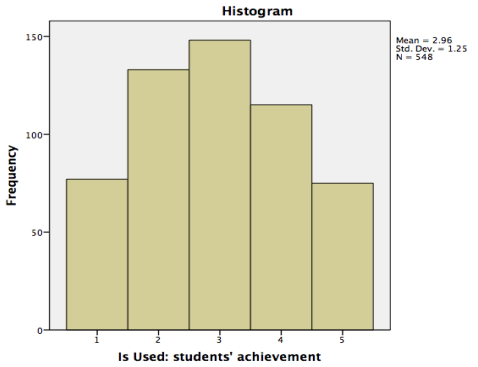
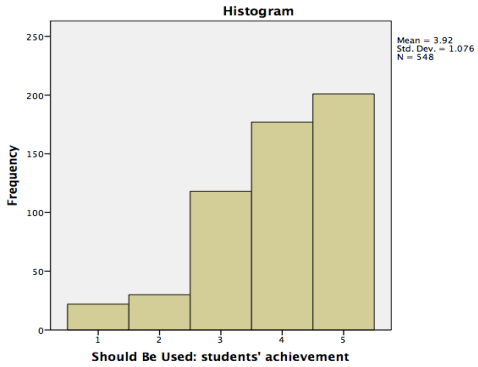
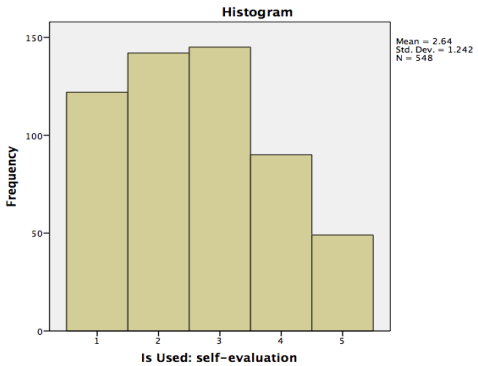
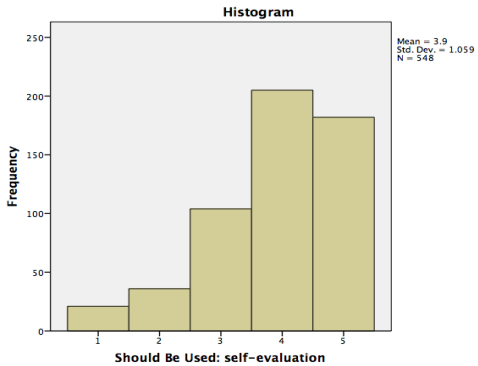
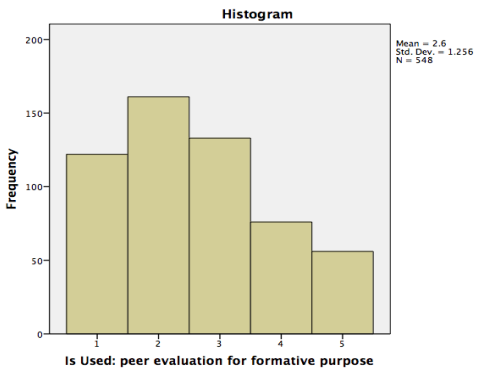
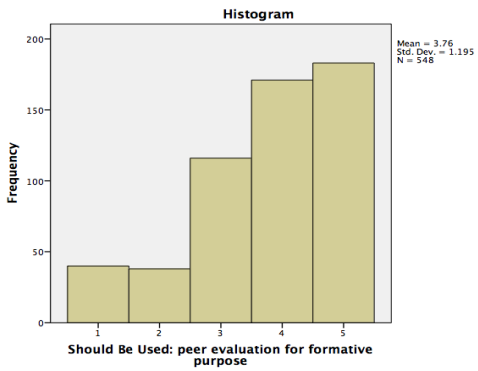
* الرجاء تحضير وتحديد الجوانب الغامضة او الاسئلة المتعلق بالنظام البديل او اي نقاط تحتاج الى توضيح اكثر حتى يتسنى لنا مناقشتها جمعيا في الحلقة النقاشية الاولى / او المقابلة للموجهين والمدراء

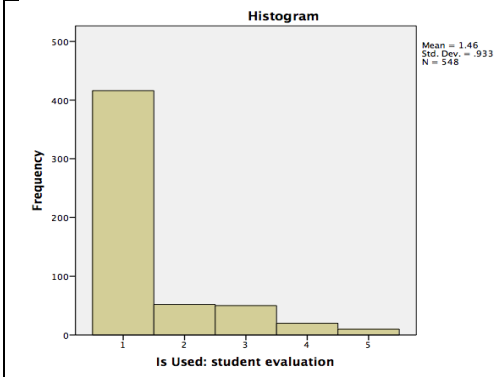
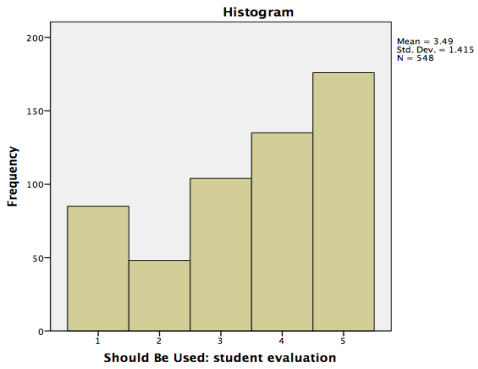
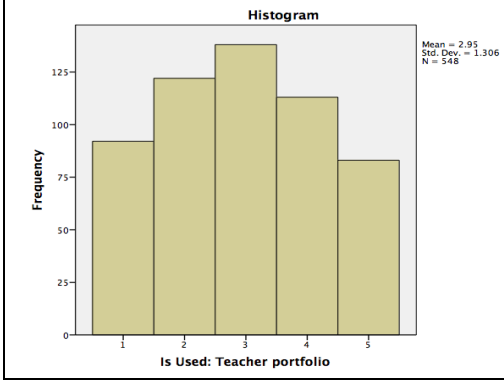
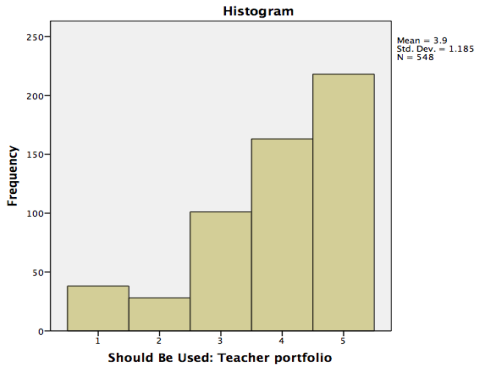
اعداد: طلال سعد المطيري
باحث دكتوراه
جامعة درم / المملكة المتحدة

Appendix (21): Histogram of the Ceiling and Flooring Effect for the Purposes of Teacher Evaluation

Item 1 A	Item 1 B
	
Item 2 A	**Item 2 B**
	
Item 3 A	**Item 3 B**
	

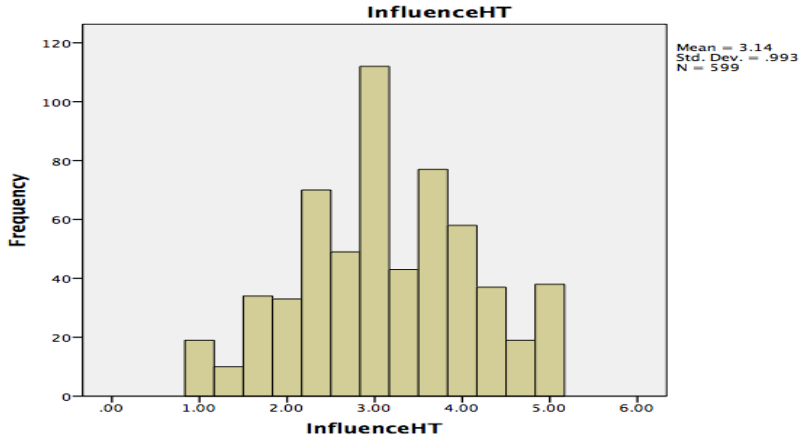
Appendix (22): Histogram of the Ceiling and Flooring Effect for the Tools of Teacher Evaluation

Item 4 A (Observation)	Item 4 B (Observation)
	
Item 5 A (Students' achievement)	**Item 5 B (Students' achievement)**
	
Item 6 A (self-evaluation)	**Item 6 B (self-evaluation)**
	
Item 7 A (peer evaluation)	**Item 7 B (peer evaluation)**
	

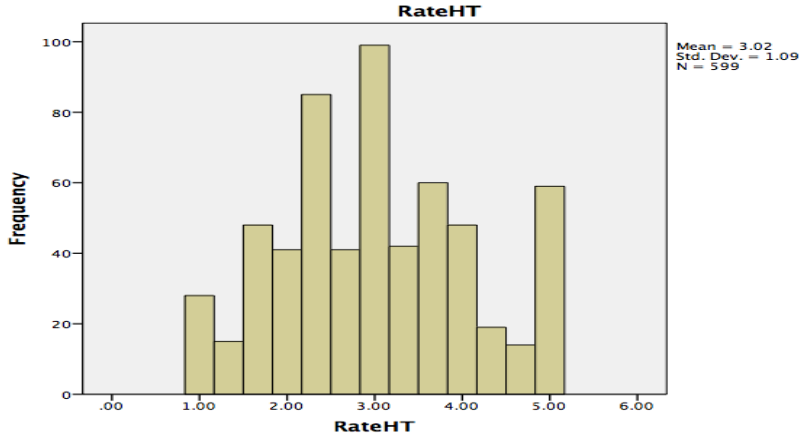
Item 8 A (student evaluation)	Item 8 B (student evaluation)
	
Item 9 A (teacher portfolio)	**Item 9 B (teacher portfolio)**
	

Appendix (23): Histogram of the Ceiling and Flooring Effect for the Involvement of Evaluators

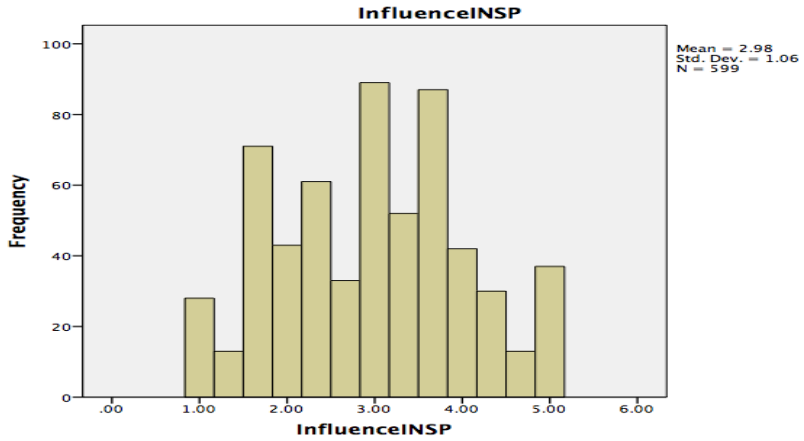
Heads Teacher Role:



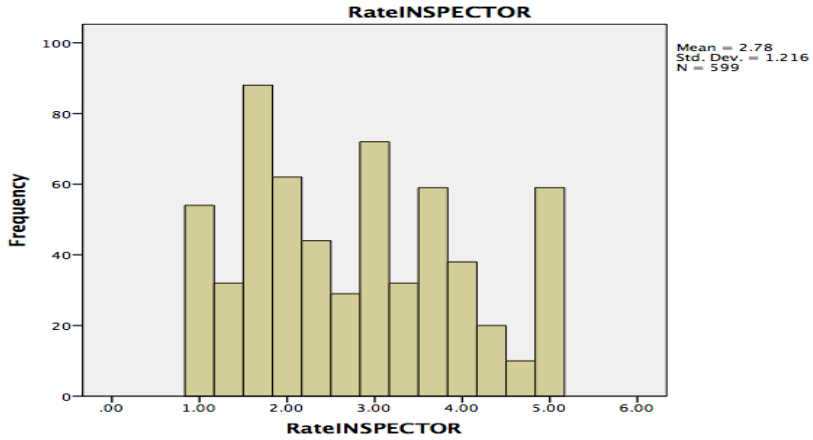
Rating the Value of Heads Teacher:



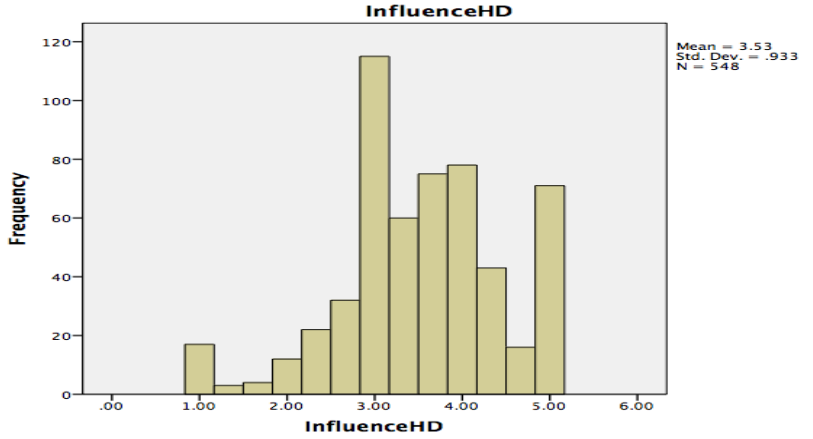
Inspectors Role:



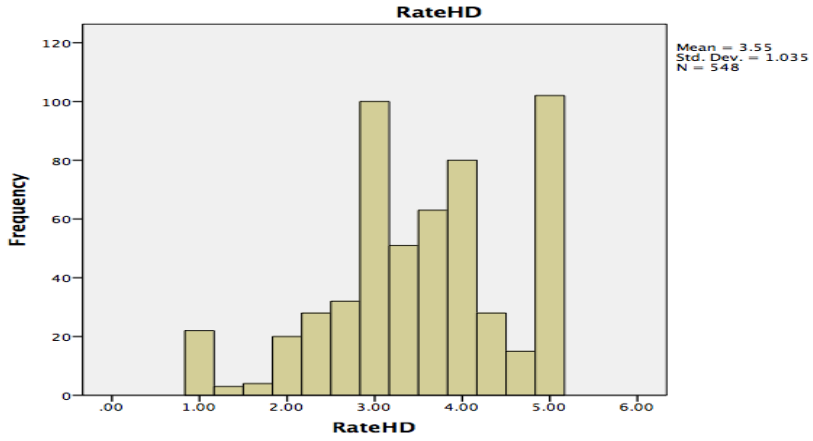
Rating the Value of Inspectors:



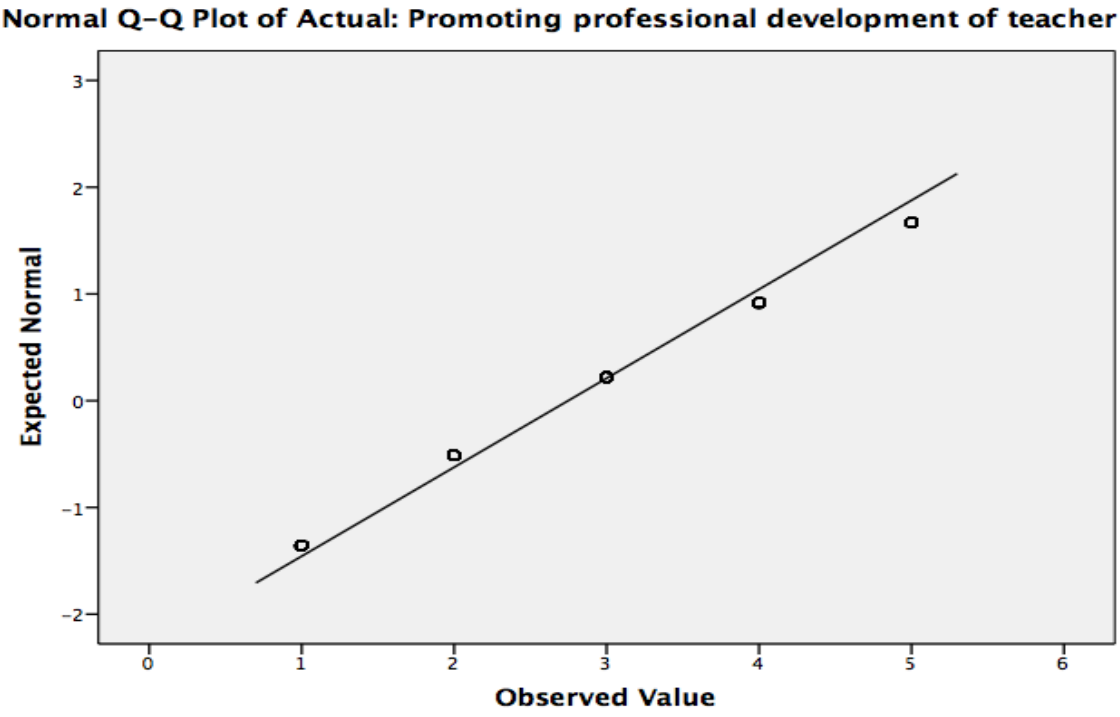
Heads Departments Role:

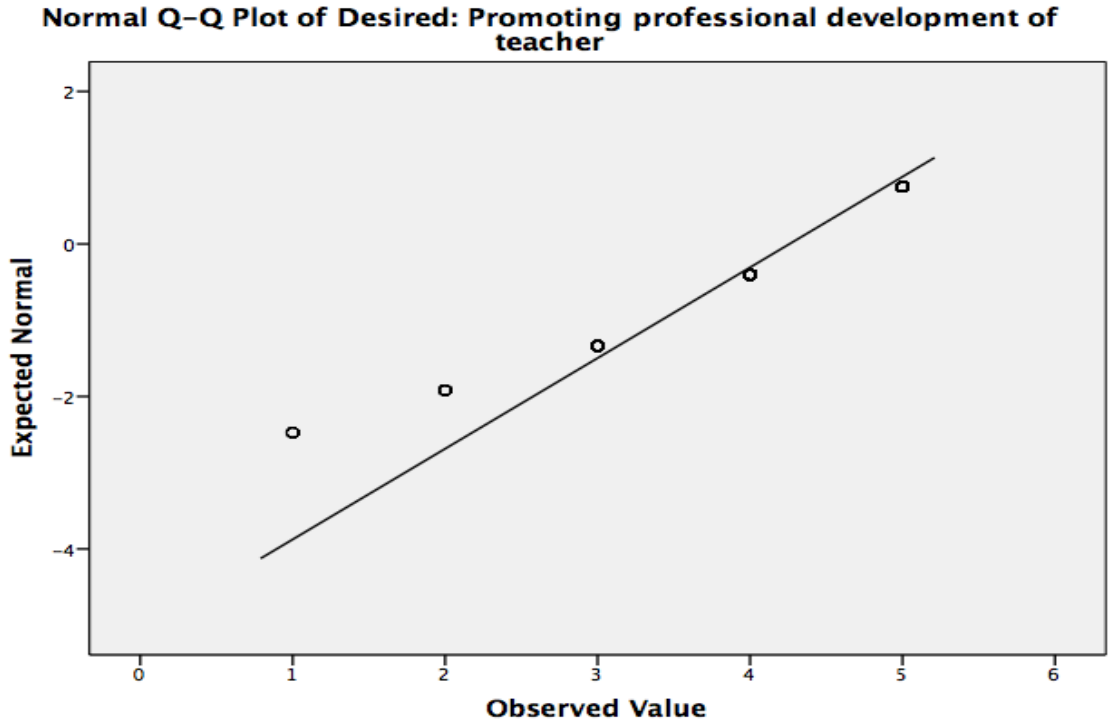


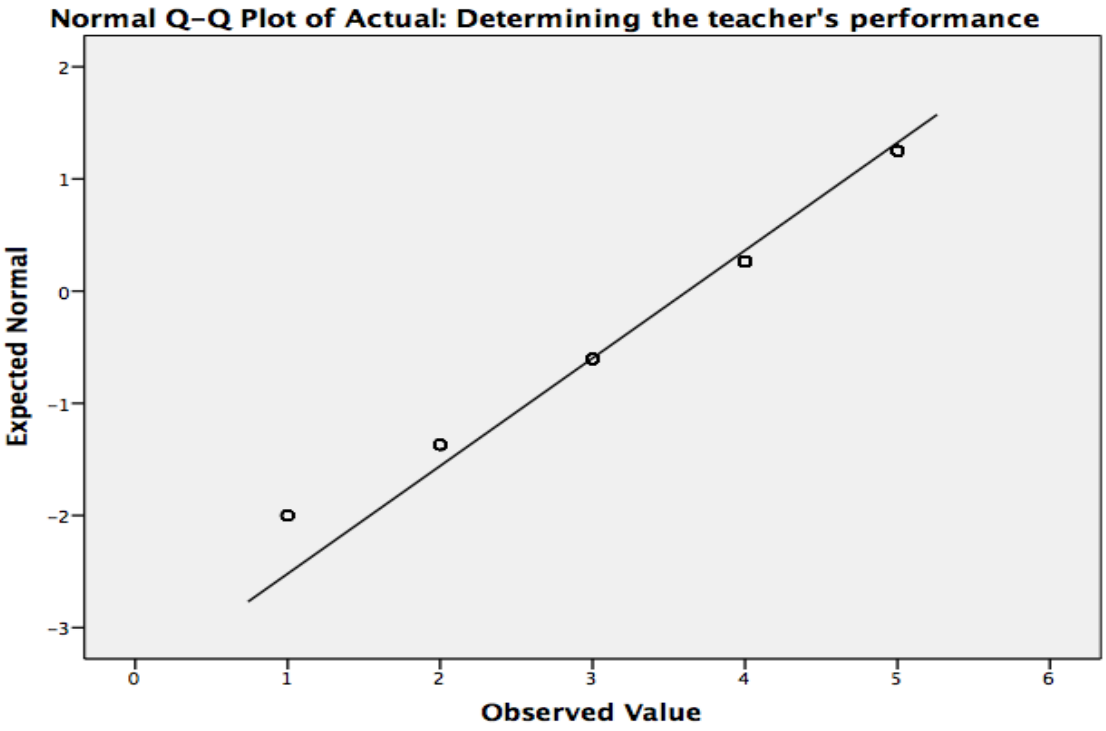
Rating the Value of Heads Departments:

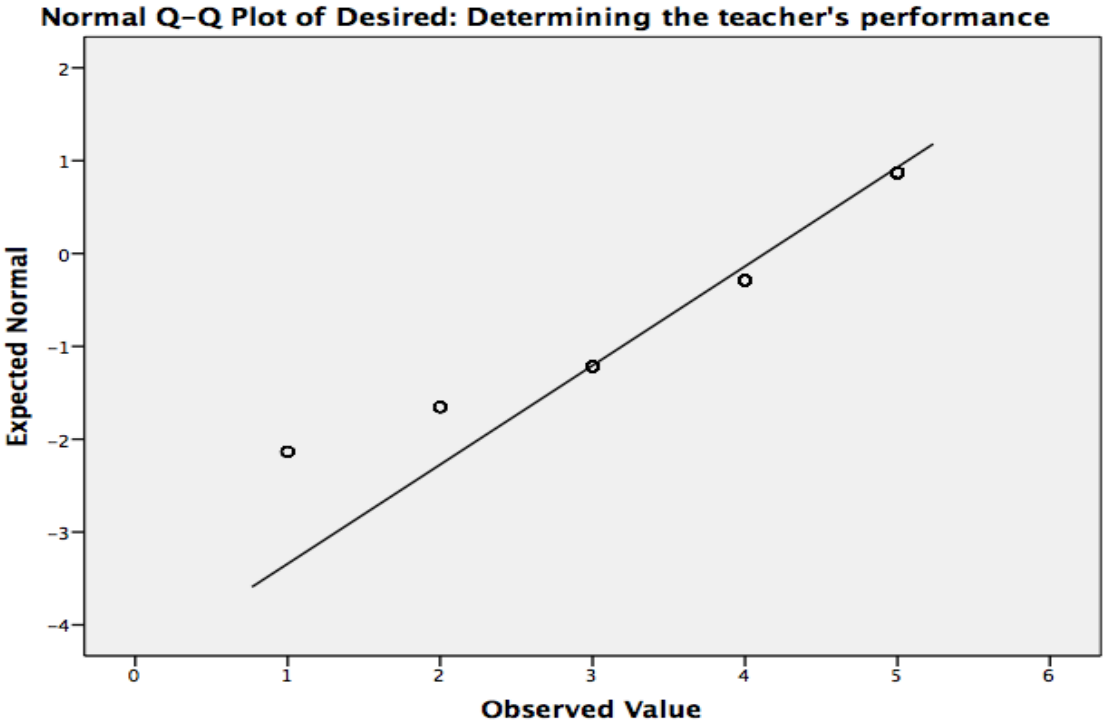


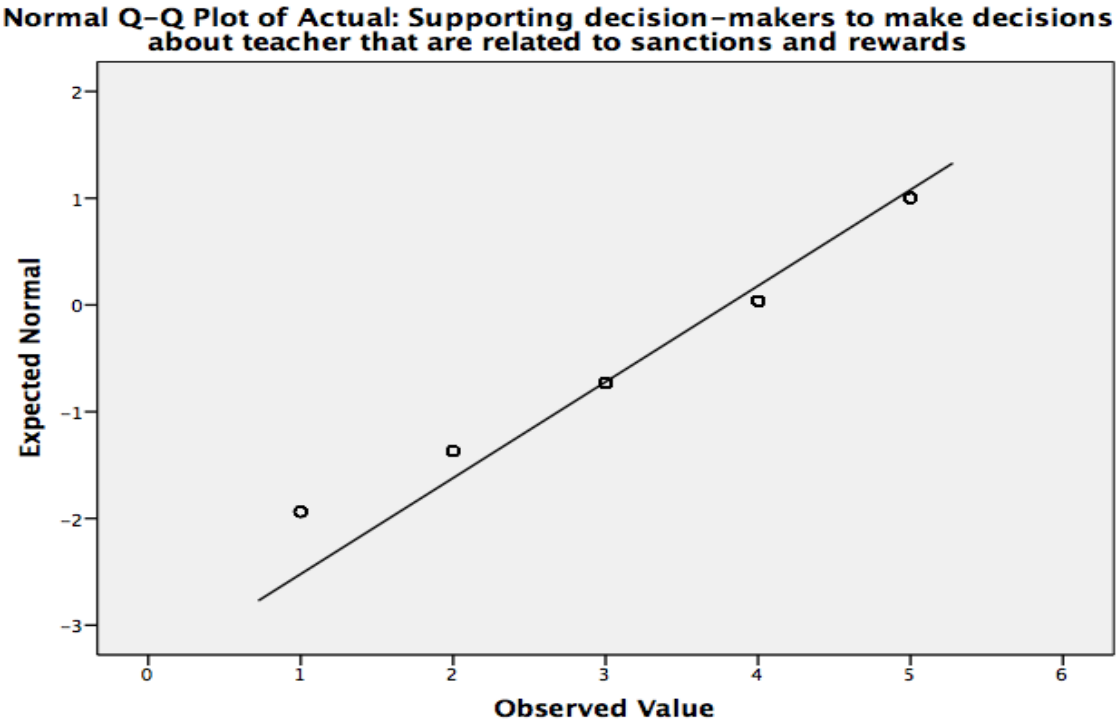
Appendix (24): Tests of Normality for the Purposes of Teacher Evaluation (Actual/Desired)

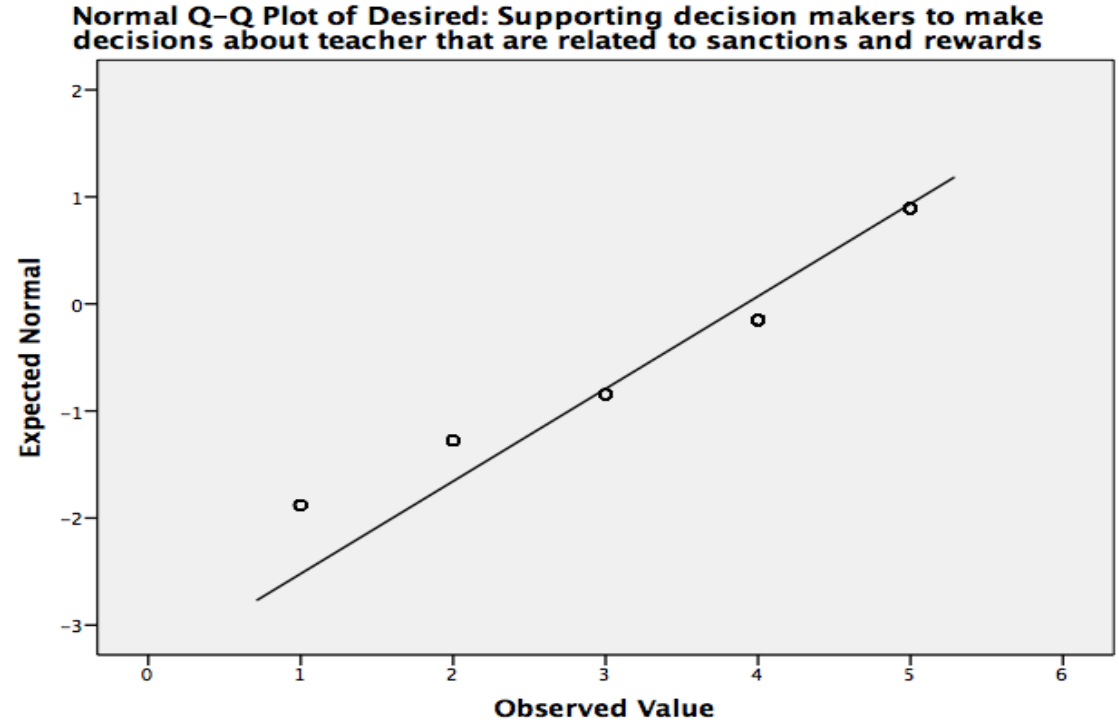




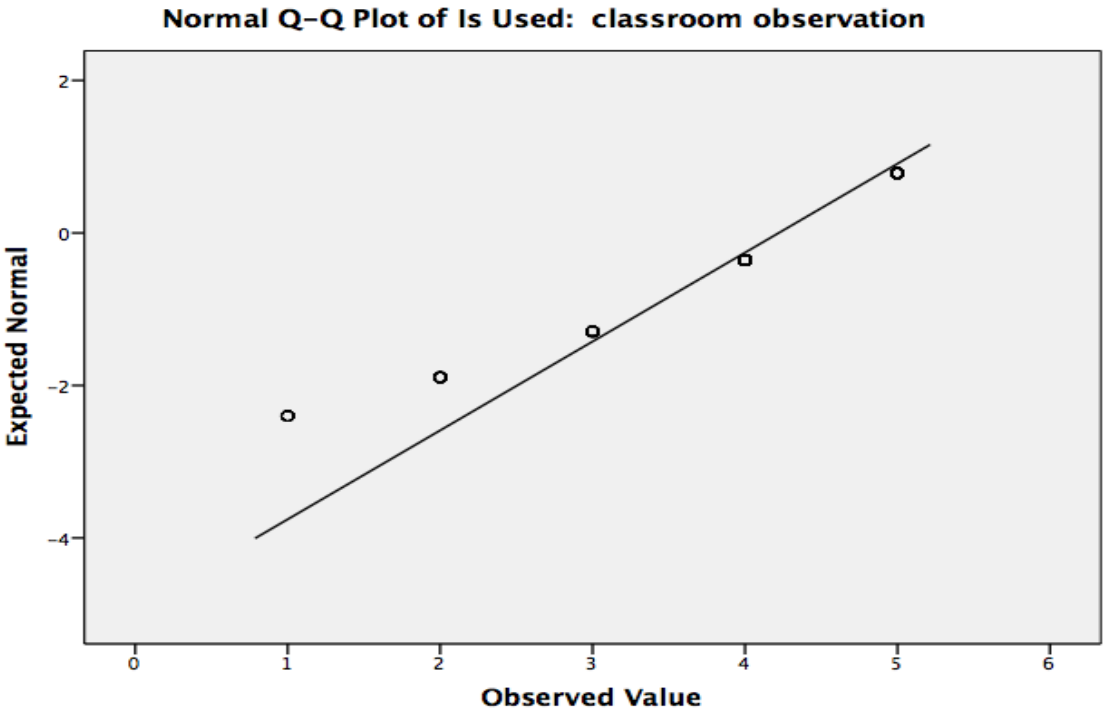

Normal Q-Q Plot of Actual: Determining the teacher's performance
Expected Normal
Observed Value
2
1
0
-1
-2
-3
0
1
2
3
4
5
6

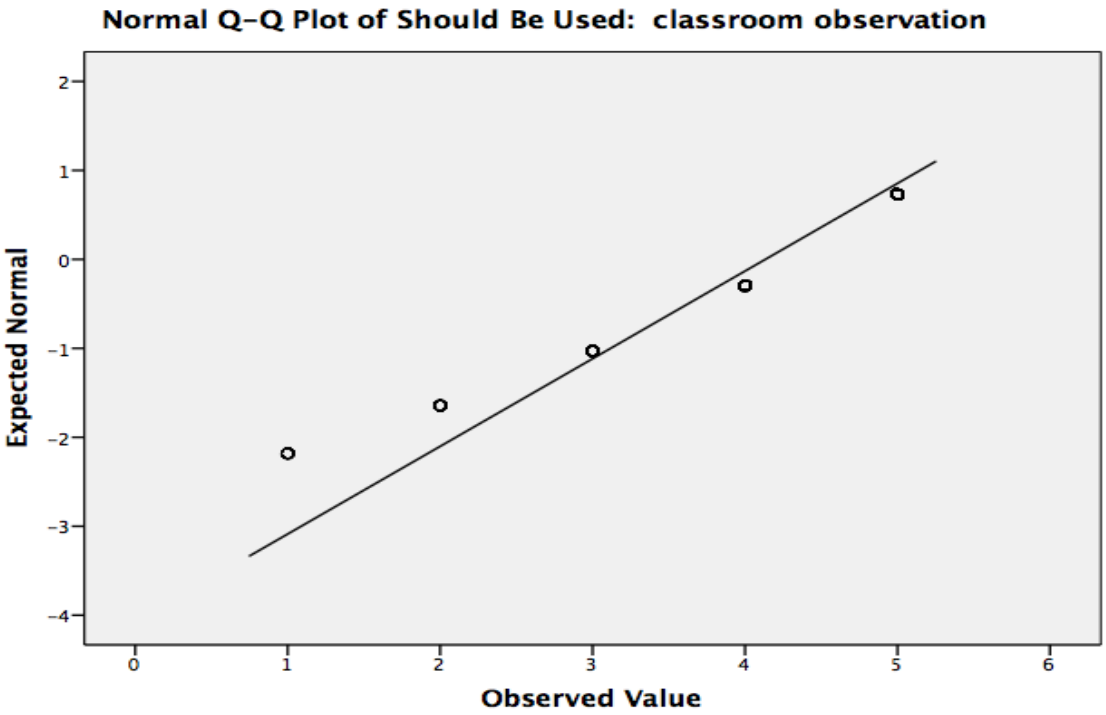

Normal Q-Q Plot of Desired: Determining the teacher's performance
Expected Normal
Observed Value
2
1
0
-1
-2
-3
-4
0
1
2
3
4
5
6

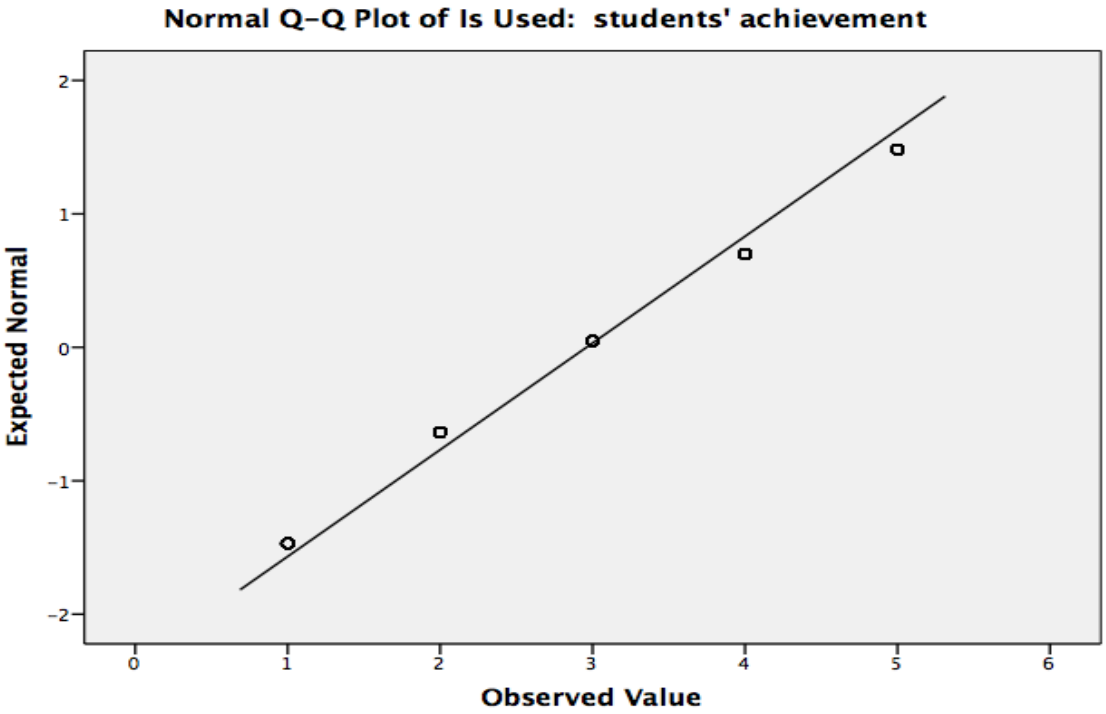

Normal Q-Q Plot of Actual: Supporting decision-makers to make decisions about teacher that are related to sanctions and rewards
Expected Normal
Observed Value
2
1
0
-1
-2
-3
0
1
2
3
4
5
6

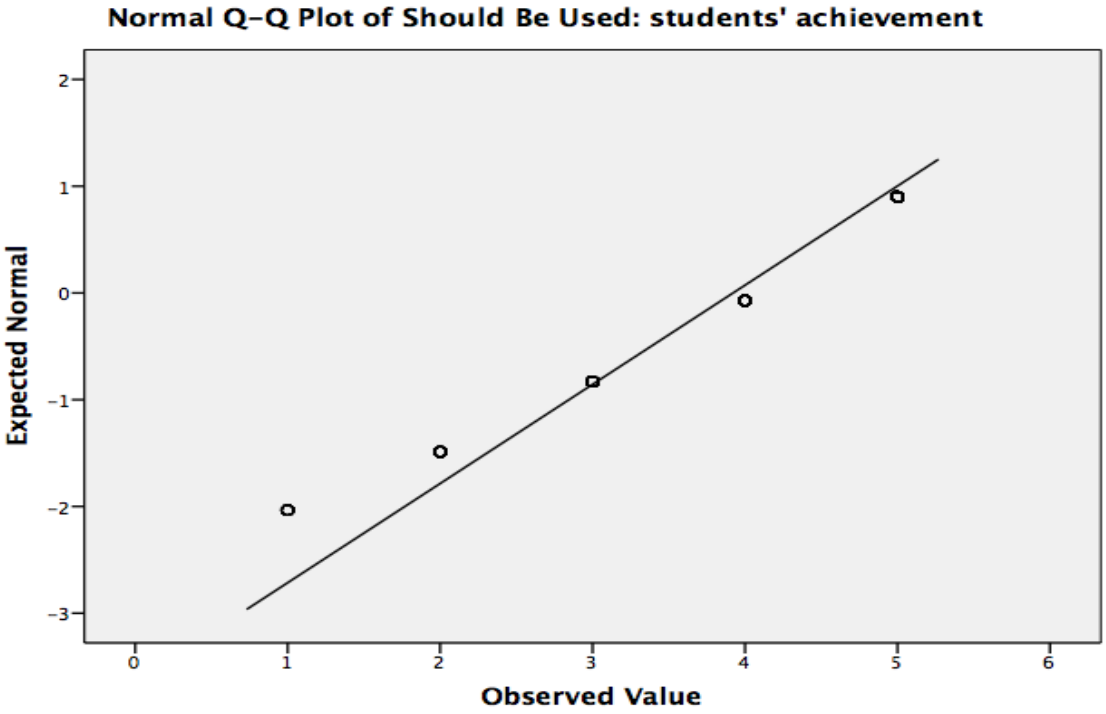

Normal Q-Q Plot of Desired: Supporting decision makers to make decisions about teacher that are related to sanctions and rewards
Expected Normal
Observed Value
2
1
0
-1
-2
-3
0
1
2
3
4
5
6

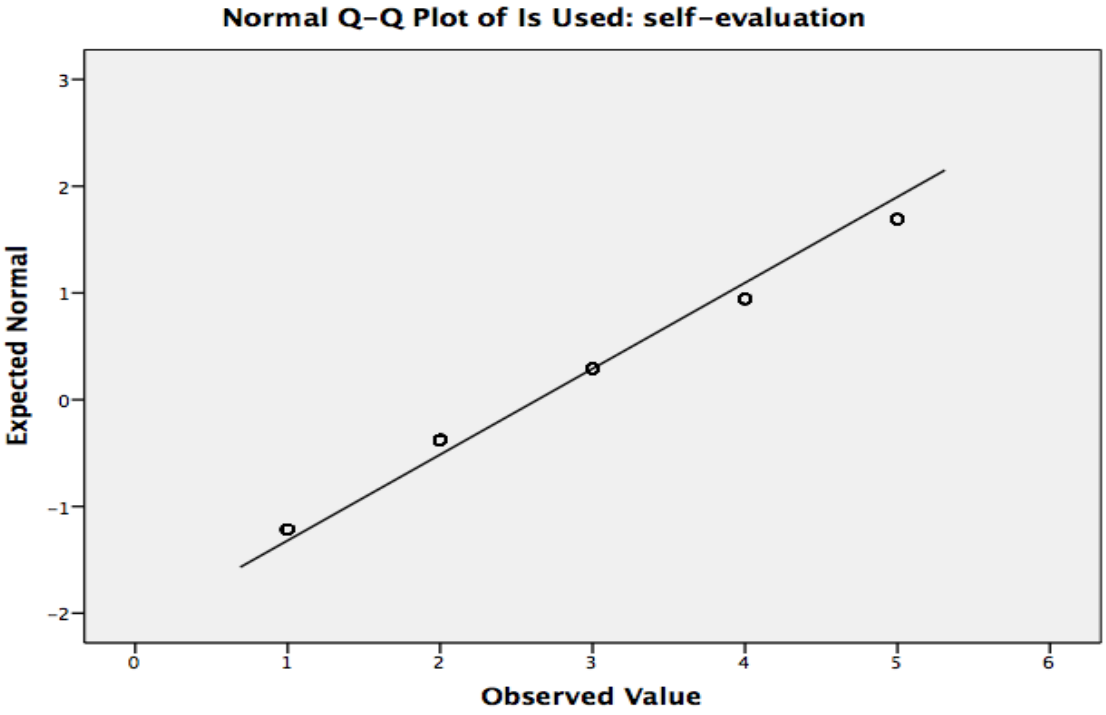
Appendix (25): Tests of Normality for the Tools of Teacher Evaluation (Is used/Should be used)

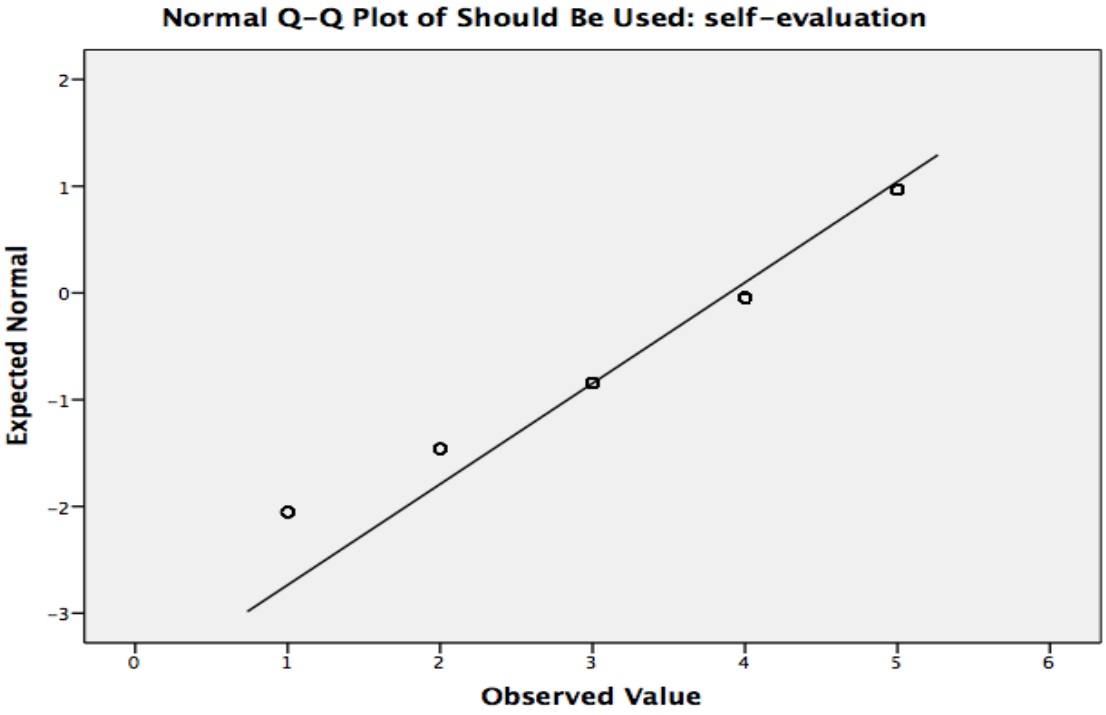


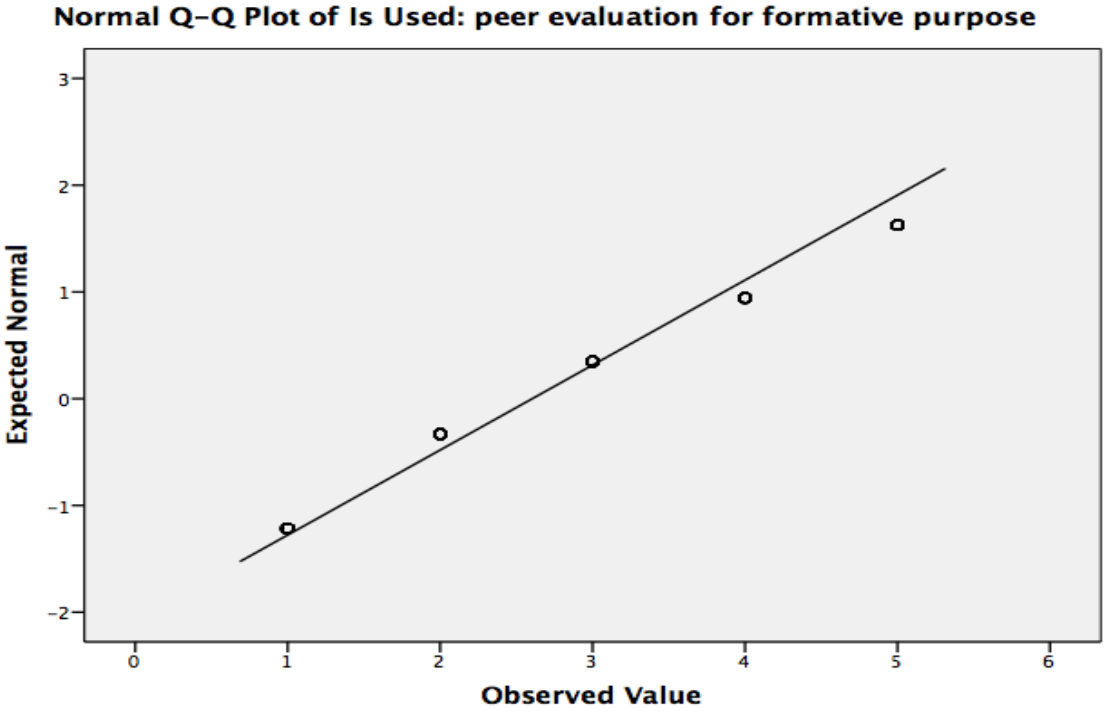


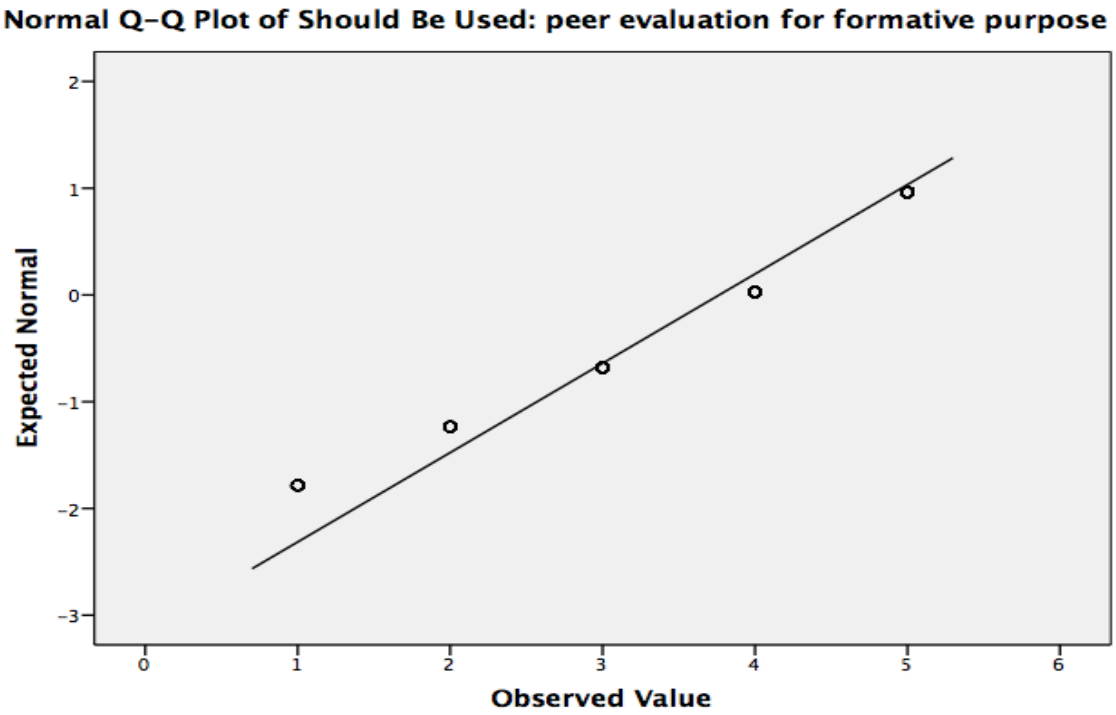

Normal Q-Q Plot of Is Used: students' achievement
Expected Normal
Observed Value
2
1
0
-1
-2
0
1
2
3
4
5
6

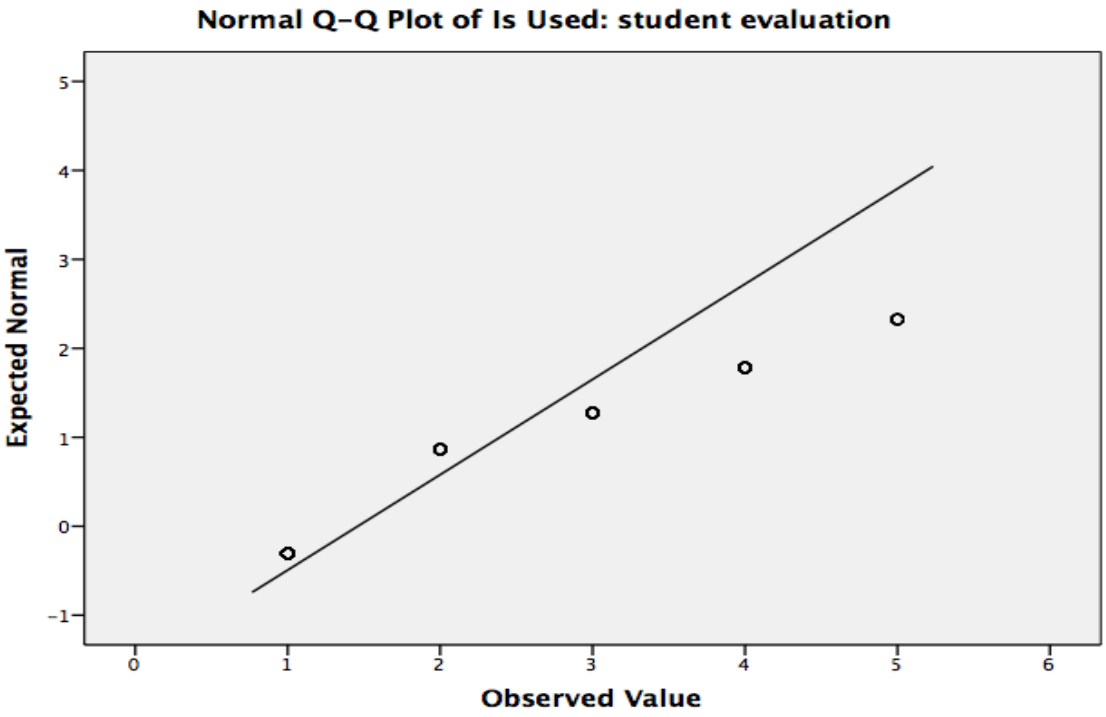

Normal Q-Q Plot of Should Be Used: students' achievement
Expected Normal
Observed Value
2
1
0
-1
-2
-3
0
1
2
3
4
5
6

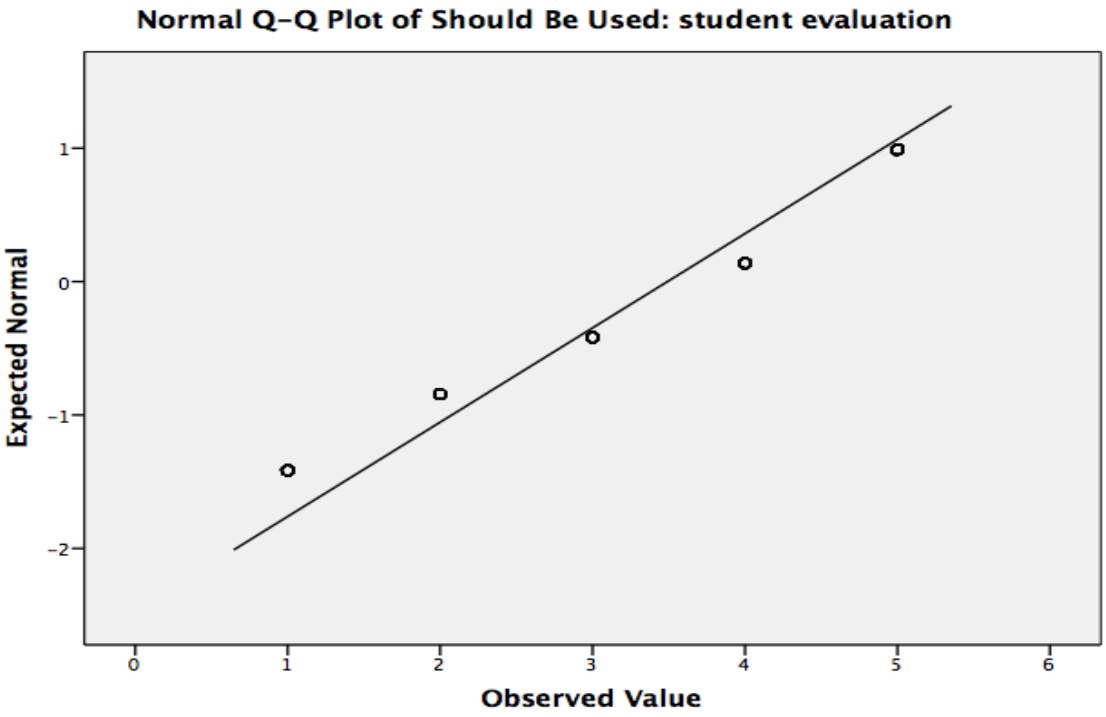

Normal Q-Q Plot of Is Used: self-evaluation
Expected Normal
Observed Value
3
2
1
0
-1
-2
0
1
2
3
4
5
6

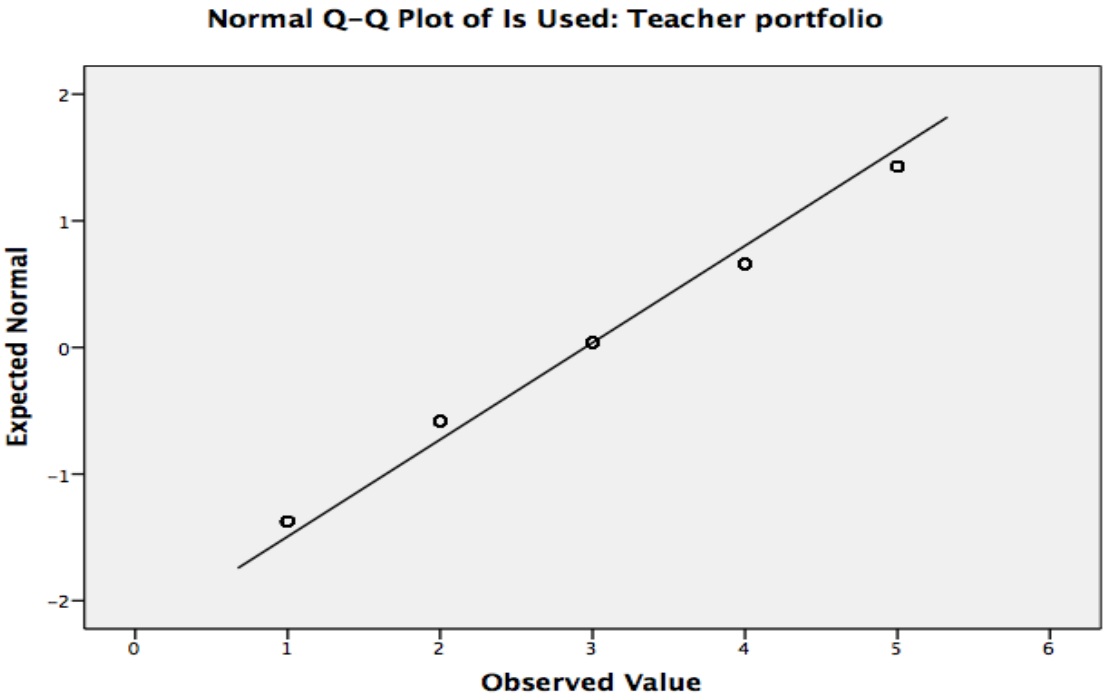

Normal Q-Q Plot of Should Be Used: self-evaluation
Expected Normal
Observed Value
2
1
0
-1
-2
-3
0
1
2
3
4
5
6

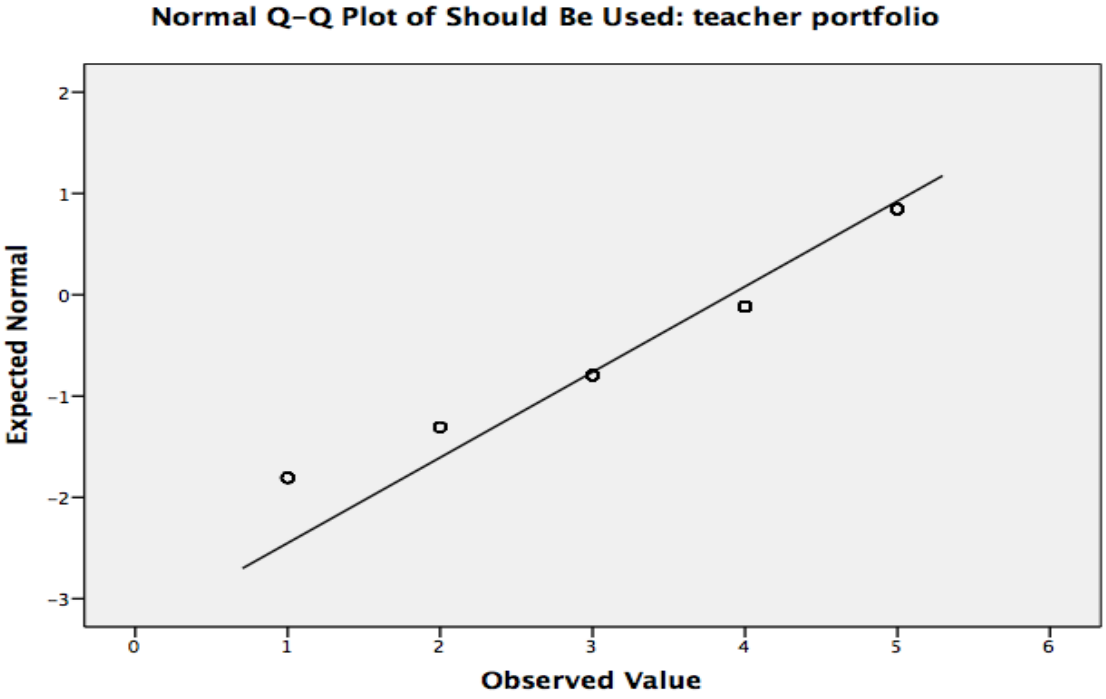

Normal Q-Q Plot of Is Used: peer evaluation for formative purpose
Expected Normal
Observed Value
3
2
1
0
-1
-2
0
1
2
3
4
5
6


Normal Q-Q Plot of Should Be Used: peer evaluation for formative purpose
Expected Normal
Observed Value
2
1
0
-1
-2
-3
0
1
2
3
4
5
6


Normal Q-Q Plot of Is Used: student evaluation
Expected Normal
Observed Value
5
4
3
2
1
0
-1
0
1
2
3
4
5
6

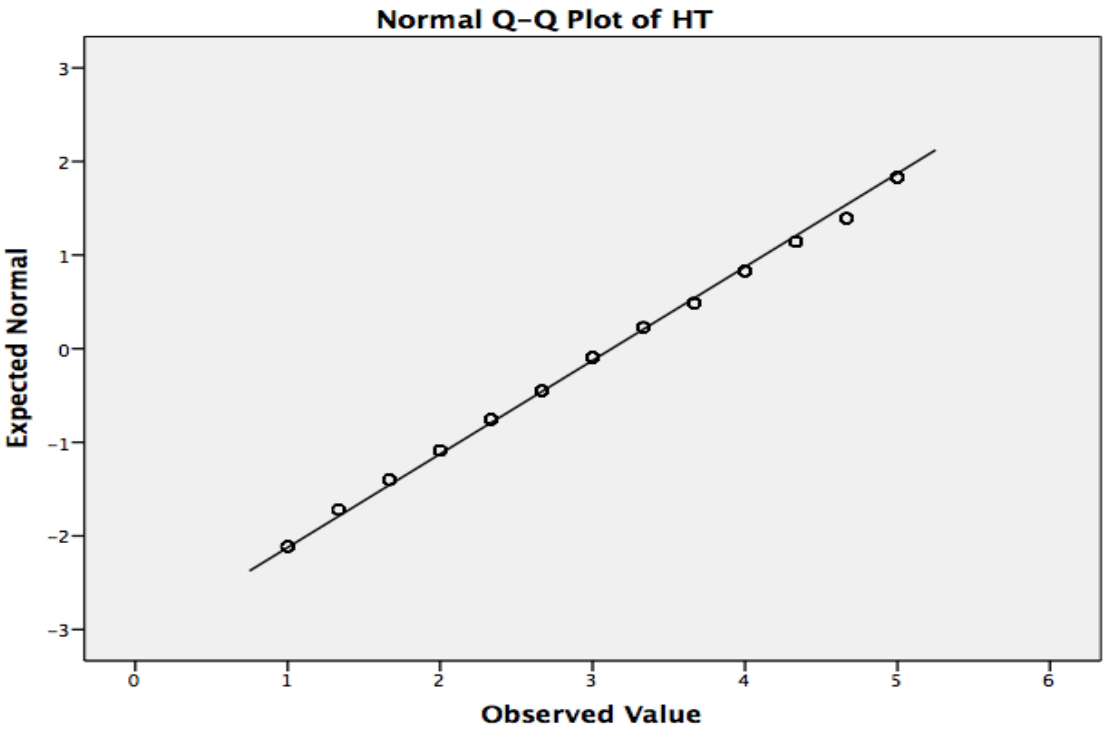

Normal Q-Q Plot of Should Be Used: student evaluation
Expected Normal
Observed Value
1
0
-1
-2
0
1
2
3
4
5
6


Normal Q-Q Plot of Is Used: Teacher portfolio
Expected Normal
Observed Value
2
1
0
-1
-2
0
1
2
3
4
5
6

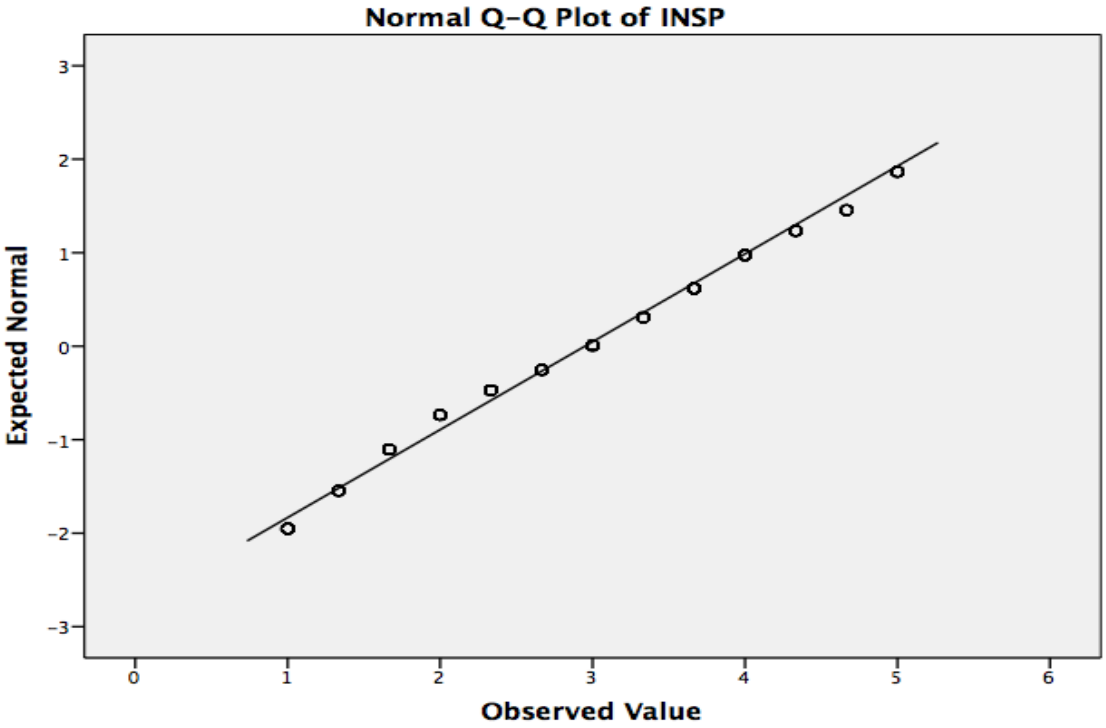

Normal Q-Q Plot of Should Be Used: teacher portfolio
Expected Normal
Observed Value
2
1
0
-1
-2
-3
0
1
2
3
4
5
6

Appendix (26): Tests of Normality for the Involvement of Evaluators

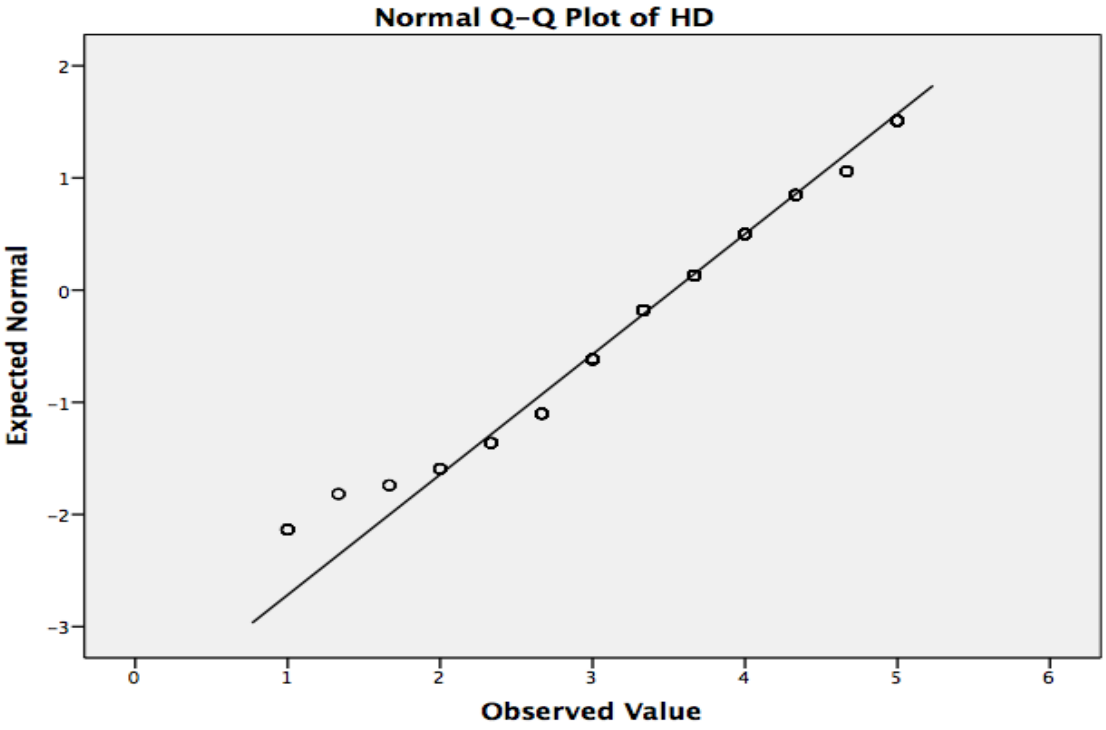
- **The Role of Head Teacher**



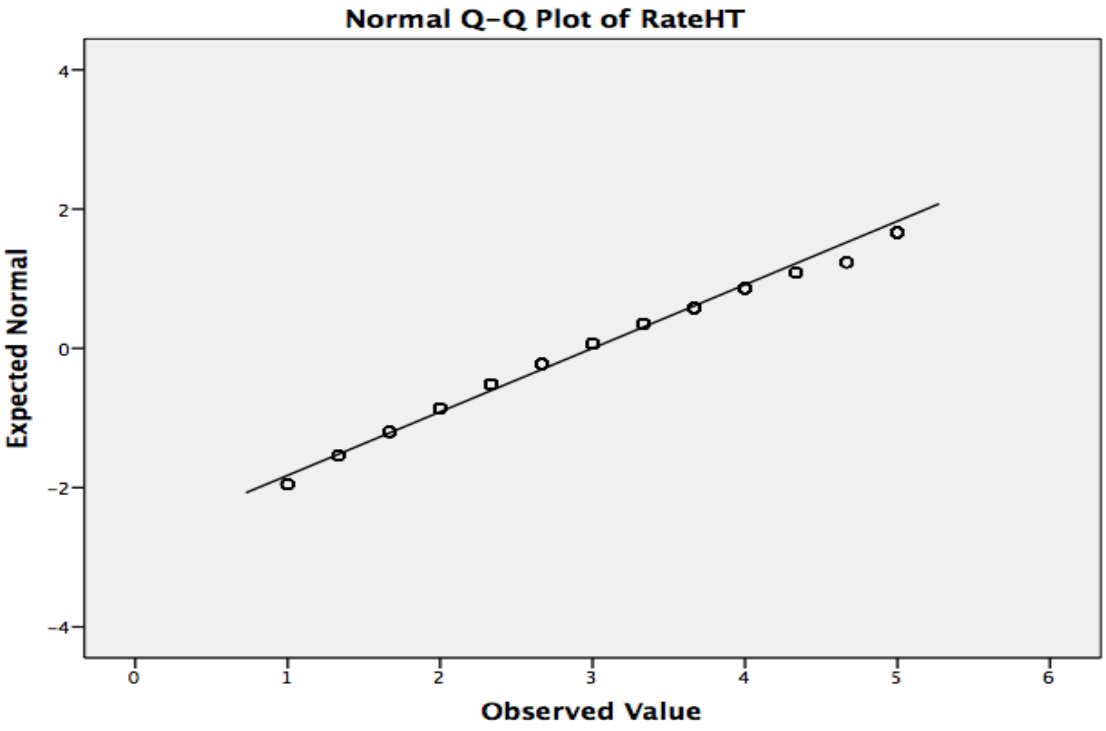
- **The Role of Inspector**



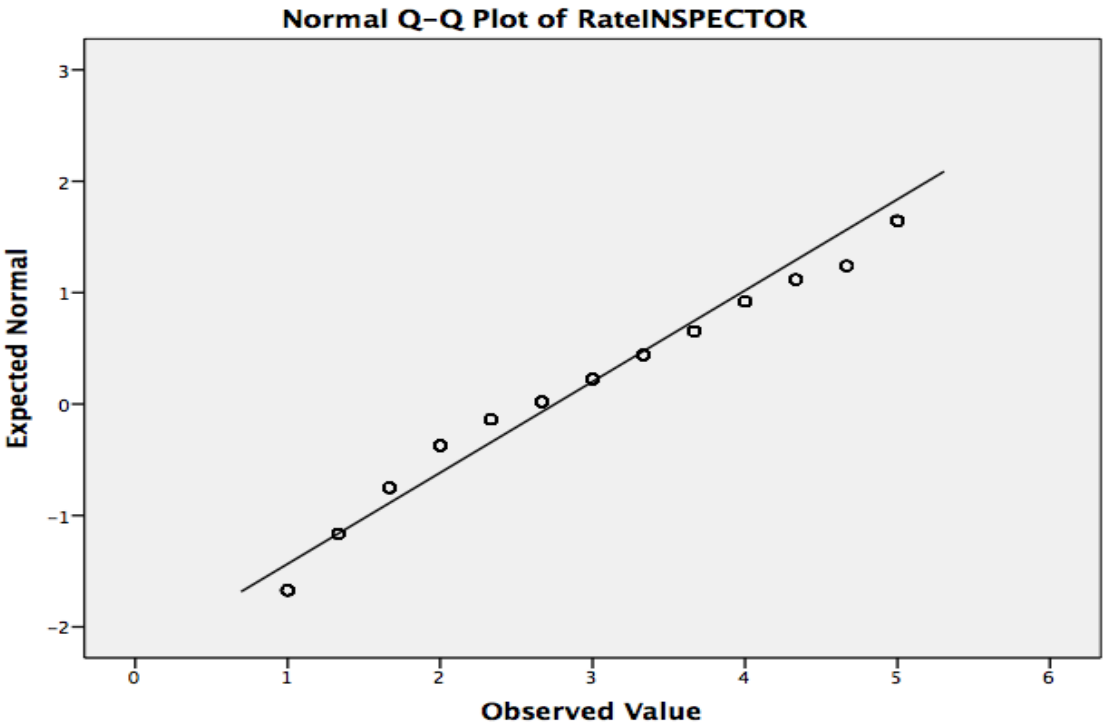
- **The Role of Head Department**



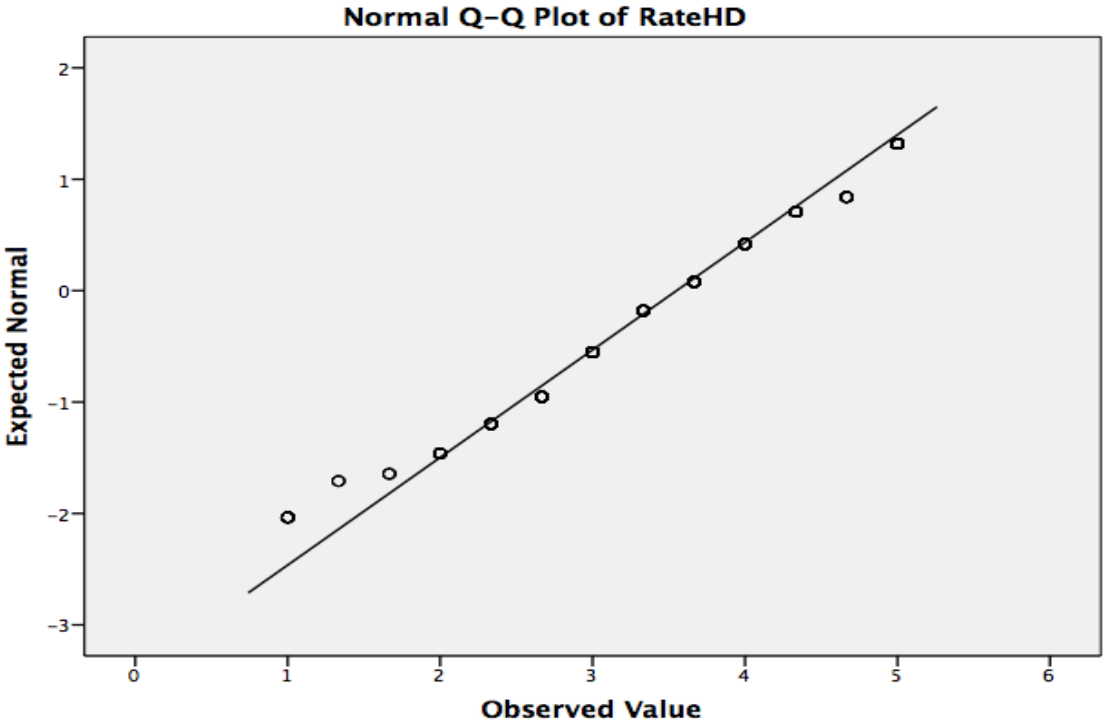
- **Rating Value [Head Teacher]**



- **Rating Value [Inspector]**

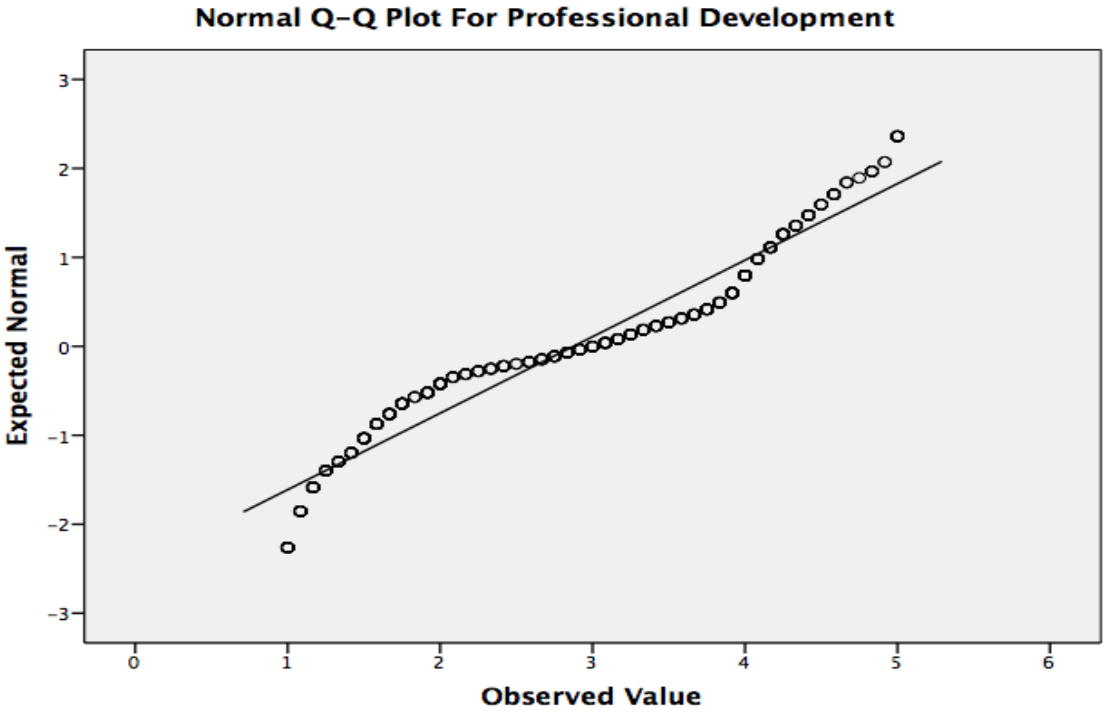


- **Rating Value [Head Department]**

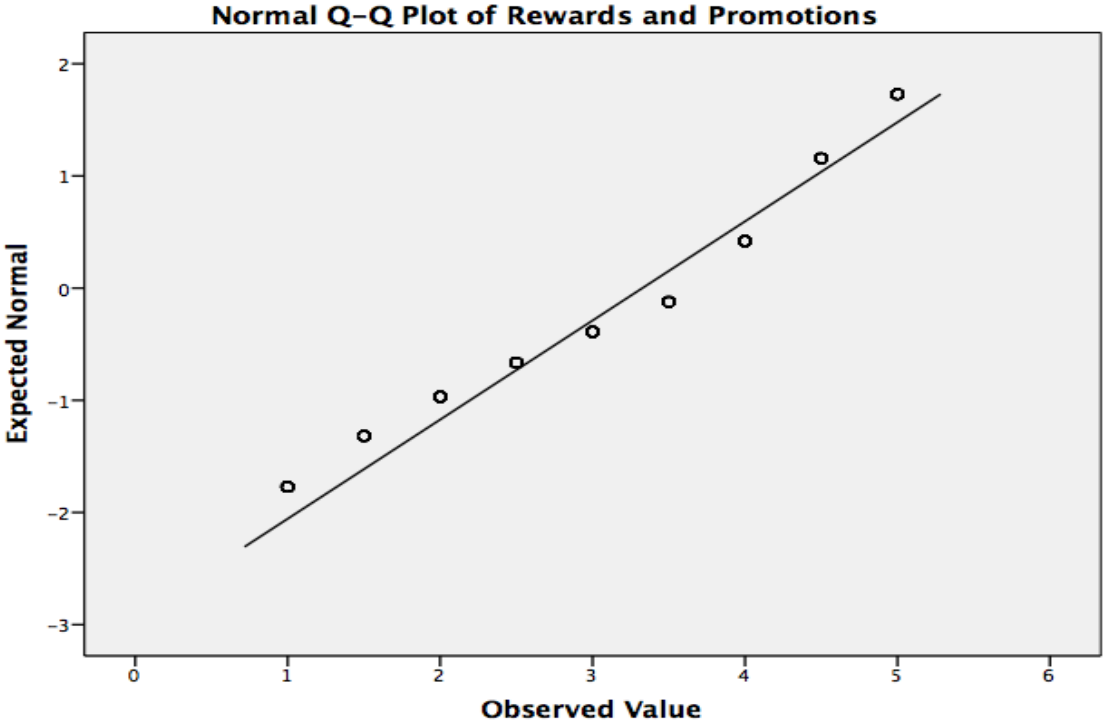


Appendix (27): Tests of Normality for the Extent to Which the System Supports Teachers

- **The Current system supports the development of performance**



- **The Current system supports the awarding of promotions and rewards**



Appendix (28): Test of Normality (significance)

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Actual: Promoting professional development of teachers	.170	548	.000	.910	548	.000
Desired: Promoting professional development of teachers	.263	548	.000	.771	548	.000
Actual: Determining the teacher's performance	.224	548	.000	.884	548	.000
Desired: Determining the teacher's performance	.288	548	.000	.767	548	.000
Actual Supporting decision-makers to make decisions about teachers that are related to sanctions or rewards	.226	548	.000	.856	548	.000
Desired: Supporting decision-makers to make decisions about teachers that are related to sanctions or rewards	.275	548	.000	.808	548	.000
Is used: classroom observation	.249	548	.000	.779	548	.000
Should be used: classroom observation	.267	548	.000	.790	548	.000
Is used: students' achievement	.162	548	.000	.912	548	.000
Should be used: student's achievements	.219	548	.000	.838	548	.000
Is used: self-evaluation	.178	548	.000	.901	548	.000
Should be used: self-evaluation	.245	548	.000	.843	548	.000
Is used: peer evaluation for formative purpose	.201	548	.000	.894	548	.000
Should be used: peer evaluation for formative purpose	.224	548	.000	.848	548	.000
Is used: student evaluation	.448	548	.000	.558	548	.000
Should be used: student evaluation	.208	548	.000	.849	548	.000
Is used: teacher portfolio	.157	548	.000	.906	548	.000
Should be used: teacher portfolio	.228	548	.000	.817	548	.000
Role of Head teacher	.107	548	.000	.974	548	.000
Role of Inspector	.091	548	.000	.966	548	.000
Role of Head of department	.119	548	.000	.947	548	.000
Rate value- Head teacher	.105	548	.000	.961	548	.000
Rate value- Inspector	.138	548	.000	.933	548	.000
Rate value- head of department	.105	548	.000	.939	548	.000
The system supports for professional development	.131	548	.000	.926	548	.000
The system supports for rewards and promotions	.233	548	.000	.907	548	.000
	a. Lilliefors Significance Correction					