Teacher Effectiveness: A Case Study Using Value-Added method to measure teacher effectiveness in one UK Curriculum School in Dubai

فعالية المعلم: دراسة تجريبية لاستخدام طريقة القيمة المضافة في علامات الطلبة لتقييم أداء المعلمين في احدى المدارس ذات المناهج البريطاني في دبي

By
Student name: Raed Bdeir

Student ID number: 2013201013

Dissertation submitted in partial fulfillment of MEd Management leadership and Policy

Faculty of Education

Dissertation Supervisor
Dr. Clifton Chadwick

July-2015

To My Wife

This research is dedicated to my beloved wife for her support; encouragement and patience with me while doing my masters and writing my dissertation.

Abstract

This study aims to identify whether value-added is a valid way of measuring a teacher’s effectiveness. The research will investigate whether effective teaching has an impact on value-added.

Results from international studies which have used value added to measure teacher effectiveness have proved inconclusive. While researchers in the United States of America (USA) found that this method was limited in measuring a teacher’s effectiveness, those in the United Kingdom (UK) found it to be a useful tool for identifying good and weak teachers. The lack of research in Dubai and the United Arab Emirates (UAE) into the use of value added as a means of determining whether or not a teacher was effective acted as a foundation for this study. The researcher selected one private school in Dubai offering the UK curriculum in which to carry out this research.

Results were obtained by a process of triangulation using a mixed method research type. A sample of male and female students from years 10 and 11, taught by a mixed gender sample of teachers was selected. The students’ end-of-year results in biology, math and physics were compared with their IGCSE exam results in these subjects. The teachers’ appraisals were also scrutinized to see if there was a link between the quality of teaching reported in classroom observations and exam results.

The research revealed that good teaching did have a positive effect on students’ outcomes, thereby confirming that value added is an appropriate tool to measure teacher effectiveness.

الخلاصة:

ركزت هذه الدراسة على التحقق من أن القيمة المضافة في علامات الطلبة تُمثِّل أداةً صحيحة لقياس فعالية المعلمين، والتحقق كذلك من وجود تأثير إيجابي للتدريس الفعال على هذه القيمة المضافة.

ولا بد من الإشارة إلى وجود دراسات دولية استخدمت طريقة القيمة المضافة في قياس فعالية المعلمين، إلا أن هذه الدراسات لم تقدم نتائج حاسمة في هذا الجانب، في حين اعتبر باحثون من الولايات المتحدة الأمريكية أن هذا الأسلوب محدود في قياس فعالية المعلمين، رأى الباحثون في المملكة المتحدة أن هذا الأسلوب يقدم أداةً مفيدة لتحديد المعلم الجيد والمعلم الضعيف. ولا ريب أن حاجتنا في دبي والإمارات العربية المتحدة إلى إجراء أبحاث ودراسات حول استخدام القيمة المضافة كأداة لتحديد فعالية المعلمين، كانت الدافع لإجراء هذه الدراسة ولقد اختار الباحث إحدى المدارس الخاصة في دبي التي تطبق منهاجاً تعليمياً بريطانياً لإجراء الدراسة عليها.

لقد تم الحصول على النتائج بتطبيق منهجية التثليث وباعتماد مصادر متعددة في البحث. وتم اختيار عينة من الطلبة البنين والبنات في السنتين الدراسيتين العاشرة والحادية عشرة يتولى تدريسهم معلمين ومعلمات. وتمت مقارنة نتائجهم في امتحانات نهاية العام الدراسي مع نتائجهم في امتحانات الشهادة الثانوية العامة

الدولية (IGCSE) في مواد علم الأحياء والرياضيات والفيزياء، والتدقيق في عمليات تقييم أداء هؤلاء المعلمين للتحقق من وجود أية علاقة بين جودة تدريسهم في الغرف الدراسية ونتائج الامتحانات.

لقد كشفت هذه الدراسة عن وجود تأثير إيجابي للتدريس الجيد على مخرجات الطلبة، مما يؤكد أن طريقة القيمة المضافة تشكّل أداةً ملائمة لقياس مدى فعالية المعلمين.

Table of Contents

Tables

List of Tables

Chapter 1 Introduction

1.1 Overview

McCaffrey et al (2003) defined the Value-Added Method (VAM) as a set of statistical methods that use students' test score data over the years to estimate the effect of individual schools or teachers. In addition to, Piro, Wiemers and Shutt (2010), Dossett et al (2003) and Hammond et al (2012) suggested that an increase in results in a reliable and valid test where the student is linked to a specific teacher and is unaffected by other factors, can be considered to be a value added factor.

Value-Added Method was used over the past 20 years as a measure of a teacher's effect on students' overall results. Schools use different tools to evaluate teachers' effectiveness prior to the annual renewal of contracts. These include: Students' questionnaires and classroom observations by school leaders.

There has been a divergence of opinion as to the validity of this tool of enquiry in the US where this method was limited in measuring teacher's effectiveness. On the other hand, UK researchers found this method to be successful in measuring teacher's effectiveness and could distinguish between good and weak teachers.

1.2 Statement of the problem

To use the value added method or not, as an evaluation tool for teachers' effectiveness is related to the two following viewpoints namely the American one and the British one. The Americans conducted their studies in USA schools and concluded that the Value-Added Method had different constraints. Goe & Croft (2009) stated these reasons;

- Students' results were not only attributed to the 'teachers' effect; the value-added number provided a summation of the different factors that affected students' higher achievement.
- There was a strong and consistent correlation between how teachers teach in classrooms and value-added factor.
- Observing teachers in classrooms was not related to better teacher effectiveness.
- Finally, value added cannot be easily calculated for teachers in subjects such as Art, Music and Physical education, nor for lower elementary teachers who are teaching more than one subject to the same class.

On the other hand, the UK researchers' viewpoint such as that of Slater et al (2012), mentioned that students' higher attainment with an effective teacher is greater than with a weaker one. However, this is still difficult to quantify because of the data requirements related to collection of data over the year. This research will study the effect of a Value-Added factor measured for the same group of students taught by the same teacher for different subjects from year 10 and year 11 in a private school with a UK curriculum in Dubai. It will compare the gain in students' final results with the school's own evaluation of those teachers for each subject and each year level.

1.3 Background of the research

UAE is a country with people living in it from at least 200 countries all over the world. Dubai has at least 143 private schools offering 13 different curricula to more than 251,332 students. The teachers in these schools come from different nationalities, professional backgrounds and have different teaching experiences. This therefore will help educational policy makers to raise challenge to ensure that in this multicultural environment of schools to have at least higher outcomes of students. According to the Dubai School Inspection Bureau (DSIB) key findings report for 2012-2013, fifty one out of 143 schools offer the UK curriculum to almost 69,666 students. The percentage of these schools graded ‘good’ or ‘outstanding’ in DSIB inspections has risen from 42% in 2008-2009 to 68% in 2012-2013. One of the key strengths of these schools was the high attainment in the different core subjects such as mathematics and science (physics, chemistry and biology). In 2012-2013, seventy seven per cent of the teaching in these schools was ‘good’ or ‘outstanding’. The question is, how much of this increase in overall school performance is related to good teaching? The researcher who has extensive experience of school inspection in Dubai found that many schools were using different methods to evaluate teachers’ performance, but none was aware of how to use the Value-Added Method as a tool to measure teachers’ effectiveness.

1.4 Research questions

In this case study the researcher will investigate the importance of using Value-Added Method as a tool to measure a teacher’s effectiveness in a UK curriculum school for a teacher teaching biology and maths for year 10 and biology, maths and physics for year 11. This study will attempt to answer the following questions:

1- Could the positive or negative differences of students’ results in Year 10 and 11 in maths, physics and biology in IGCSE exams when compared with end of year results be related to a measure of effective teaching? Hence could it be a measure of teacher effectiveness?
2- Is there a difference between male and female students overall performance when their teachers are a male or a female?
3- Is there a difference between the value-added method outcome and the teachers’ appraisal performed by the senior leaders?

1.5 The significance of the research

Over the past 20 years Value-Added Method was a measure of teacher's effectiveness. It was seen as an effective method for measuring teacher's impact on students' overall results, but the important thing here is how to use this method. This may cause two problems, namely that

- The teacher may be evaluated as being ineffective and score below average, which means that any actions towards those teachers have to be considered carefully.
- Value-Added Method, like all other performance methods has a degree of error which could indicate that interpretations of a teacher's evaluation results may not be accurate. If a teacher is below average through a Value-Added Method result, he/she actually may actually be an effective practitioner and vice-verse as mentioned in Goldhaber and Hanson (2008); McCaffrey, Sass, and Lockwood (2008).

This study is important because it is the first of its kind in Dubai and UAE. It will investigate whether:

- private and public schools can use the value added method to measure teacher effectiveness
- there is a link between good teaching and improved test results and,
- there is a correlation between good teacher performance as demonstrated through improved test results, and the school's own teacher evaluation process

This method should serve as another important tool for the school's senior leadership team and principal to evaluate teachers effectively. Value-Added Method could be used easily if the teacher is teaching the same group of students over the many years is teaching this group of students and their results via valid exams are raised by the end of each year.

1.6 The Structure of this paper

In this chapter the researcher has introduced the research topic and outlined the problem, background, the main research questions and the significance of this research. The next chapter reviews the literature which discusses the different methods for measuring teacher's effectiveness and the importance of using the Value-Added Method as an indicator of teacher effectiveness. Furthermore, the chapter places the question under consideration into a historical perspective by using historical examples of using Value-Added Method. It mentions the difficulties and challenges of using such method in comparison to others. It relates Value-Added Method to other methods that the schools could use to measure teachers' effectiveness. Finally, it identifies the importance of using this method in schools in Dubai to measure the effectiveness of teacher performance. Chapter 3 describes the methodology adopted in this research. It explains the research design and describes the methods of data collection and analysis. The findings of this research are presented in Chapter 4 and comparisons are made between these research results and those of similar researches. Chapter 5 reflects on recommendations that can be taken into consideration when implementing such a method and gives different answers to the different research questions that were asked in the beginning of this research.

Chapter 2 Literature Review

2.1 Overview

Teachers are the single most important factor in the success of students and hence the whole school society. The teacher's importance in the educational process is without question. The whole quality of education delivered is a function of the teacher's degree of effectiveness Miller et al. (1989). Teacher effectiveness is the power to realize objectives agreed for teachers' work especially the work concerned with enabling students to learn. Hence, the effectiveness education depends on teacher's effectiveness Fatima and Nasreen (2011).

Evaluating effective teaching and its subsequent impact on students' progress can be complex and controversial. Educational effectiveness is a term that was developed to provide a more contained definition than terms of 'good' or 'quality' education. Effective teaching requires criteria for effectiveness. These criteria refer to the objectives of education in general and of teaching in particular. Campbell et al (2004) stated that a teacher is effective if he/she can accomplish the planned goals and assigned tasks in accordance with school goals. The same holds for teaching. According to Creamers (1999), even when we expect that schools can contribute to more than academic outcomes, and teaching is more than instruction, effective instruction remains an important component of its teachers. Teacher effectiveness is generally referred to in term of the focus on students' outcomes and the teacher behavior and classroom process that promote them. Since the teaching has extended beyond the instructional or pedagogical role in the classroom. The teacher may be facilitating his/her colleagues' teaching, engaging in broader leadership roles in the school, and thereby enhancing the quality of his/her teaching through his/her participation in professional development programmes.

Darling-Hammond et al (2010), mentioned that analysis of students' progress or learning gains measured in achievement tests can be used to produce value-added indicators of teacher effectiveness. However, these can provide only a partial source of evidence if the achievement tests do not reflect the wider goals and outcomes of education. Olivia et al (2009) presented a five-point definition of teacher effectiveness which consists of the following attributes. A teacher:

- sets a high level of outcomes of all students;
- contributes to a positive student's academic and personal development;
- uses different resources in classrooms;
- develops a positive classroom atmosphere ; and
- provides opportunities for collaboration with other teachers, administrators and parents.

2.1 Methods of evaluating teacher effectiveness

Goe et al. (2008) described the different methods used to evaluate teachers and to provide information about what makes teachers effective. These methods are:

- classroom observation
- principal evaluations
- analysis of classroom artifacts
- teaching portfolios
- teacher self-reports
- student surveys and finally
- teacher effectiveness as measured by value-added method

1- Classroom observation is used to measure observable classroom processes, including specific teacher practice, aspects of instructions and interactions between teachers and students. This can also measure broad aspects of teaching or subject-specific or content-specific aspects of practice. Some of the research done shows that this method could be linked to students' achievement. This was mentioned in Gallagher (2004), Good, Grouws and Ebmeier (1983), Hamre and Pianta (2005) and Lutz, Guthrie and Davis (2006). Research and validity findings are highly dependent on the instrument used, sampling procedures and training of observers. The main advantages of this method are that;

- it provides information about classroom behaviors and activities;
- it is considered a fair and direct measure by stakeholders. Depending on the protocol it can be used in various subjects, grades and contexts and
- finally it can provide information that is useful for the formative and summative purposes.

However the disadvantages are;

- it assesses observable classroom behavior but is not as useful as assessing beliefs, feelings, intentions or out of classroom activities;
- it is expensive due to the cost of observers' time, intensive training and moderating of observers which adds to the validity.
- attention must be made to the choice of a valid and reliable protocol, training and calibrating rates.

2- Principal evaluation is generally based on classroom observations and may be structured or unstructured. It is generally used for summative purposes, linked to performance management of objectives most commonly for tenure or dismissal decisions for new teachers. However, it is used extensively to ensure consistency of approaches across the school and to evaluate whether staff are implementing school objectives as defined in the school development plan. Studies comparing subjective principal ratings to students' achievement find mixed results. Little evidence exists on the validity of evaluation as they occur in schools, but evidence exists that training for principals is limited, which would impair the validity of their evaluation. The advantages of this method can be summarized by the fact it;

- represents a useful perspective based on the principal's knowledge of school and content.
- is generally feasible and can be one useful component in a system used to make summative judgments and provide formative feedback.

However the disadvantages are that;

- evaluation instruments can be used without proper training;
- subjectivity might impair impartial evaluation and this in turn could undermine validity.
- the principal may not be qualified to evaluate teachers on measures highly specialized for certain subjects or contents.
- the principal might not be trained to evaluate the impact of teacher methodology on student's measurable learning in the classroom.

3- Instructional artifacts are a structured protocol used to analyze classroom artifacts in order to determine the quality of instruction in a classroom. This could include lesson plans, teacher assignments, assessments, scoring rubrics and student work. More research is needed to establish scoring reliability and determine the ideal amount of work to sample. There is a lack of research on the use of structured artifact analysis in practice. Advantages of this method would demonstrate that

- it is a useful measure of instructional quality if a validated protocol is used, if observers are well trained for reliability, and if assignment show sufficient variation in quality.
- it is practical and flexible because artifacts have already been created for the classroom.

The disadvantages however are that;

- this method needs more valid and reliable research.
- training knowledgeable evaluators can be costly but is necessary to ensure validity.

This method may be a promising middle ground in terms of feasibility and validity between full observation and less direct measure such as self-report.

4- Portfolio is used to determine a wide range of teaching behaviors and responsibilities. It has been used widely in teacher education programs and in assessing teacher performance. Research on the validity and reliability of this method is ongoing, and it raises the issue about consistency of stability in evaluating. Overall there is a lack of research in linking portfolios to students' achievement. The advantages of this method are that:

- it is comprehensive;
- it can measure aspects of teaching that are not readily observed in classroom.
- It can also provide a high level of credibility among stakeholders.
- it is a good tool for teacher reflection and improvement.

Nevertheless the disadvantages of this are related to the fact that;

- it is time consuming on the part of teachers and evaluators,
- portfolios are difficult to standardize;
- portfolios represent teacher's planned work but may not reflect everyday classroom activities.

5- Teachers' self-report is when teachers report what they are doing in classrooms. This may be assessed through surveys, instructional logs and interviews. They can vary widely in focus and level of detail. Research studies on this method present mixed results. Highly detailed methods of practice may be better able to capture actual teaching practices but may be harder to establish reliability or may result in very narrowly focused methods. The advantages of this method are that;

- it can measure unobservable factors that may affect teaching, such as knowledge, intentions, expectations and beliefs;
- it can provide the unique perspective of the teacher;
- it is both feasible and cost-efficient;
- it can collect large amounts of information at once.

The disadvantages as discussed by Flowers and Hancock (2003) and Blank, Porter and Smithson (2001) are;

- the reliability and validity of self-reporting;
- the fact it is not fully established and depends on the reliability of the instrument used;
- that the use and creation of a well-developed and validated instrument will decrease cost-efficiency but will increase accuracy of findings;
- the use of this method as a sole or primary method in teacher evaluation.

6- Student survey is used to gather student opinions or judgments about teaching practice as part of teacher evaluation and to provide information about teaching as it is perceived by the students. Several studies have shown that students' ratings of teachers can be useful in providing information about teaching. They may be as valid as judgments made by college students and other groups and, in some cases, may correlate with measures of students' achievement. Validity is dependent on the instrument used and its administration and is generally recommended for formative use only. The advantages of this method are;

- that it provides a perspective from students who have the most experience with teachers;

- it can provide formative information to help teachers improve practice in a way that will connect with students;
- it makes use of students who may be as capable as the adults in evaluating an accurate rating.

The disadvantages are;

- students ratings that have not been validated for use in summative assessment and therefore should not be used as a sole or primary method of teacher evaluation;
- students cannot provide information on aspects of teaching such as teacher's content knowledge, curriculum fulfillment and professional activities.

7- Value-Added Method is used to determine teachers' contributions to students via test score gains. It may also be used to determine the distribution of the 'effective' teachers by student or school characteristics. Little is known about the validity of value-added scores for identifying effective teaching though research using Value-Added methods does suggest that teachers differ markedly in their contribution to students' test score gains. However, correlating value-added scores with teacher qualifications, characteristics, or practices has yielded mixed results and few significant findings. Thus, it is obvious that teachers vary in effectiveness, but the reasons for this are not known. The advantages of this method are that;

- it provides a way to evaluate teachers' contribution to students' learning, which most methods does not;
- it requires no classroom visits because it is linked to students'/teacher's data and can be analysed at a distance.

This method is used extensively in section 5 Ofsted inspections and as a means to develop pre inspection lines of enquiry which are subsequently followed up in an inspection. This method may be useful for identifying outstanding teachers whose classrooms can serve as ‘learning labs’ as well as struggling teachers in need to support. The disadvantages with this method are that

- it is not always possible to sort out teacher effects from classroom effects;
- vertical test alignment is assumed, that is the test essentially measures the same thing from grade to grade;
- Value-Added scores are not useful for formative purposes because teachers learn nothing about how their practices contributed to student learning;
- Value-Added methods are controversial because they measure only teachers’ contributions to student achievements gains on standardized test scores.

Little et al (2009) mentioned the purposes of evaluation of teacher effectiveness using the Value-Added Method. They found the following essential components which are;

1- To find out whether grade-level or instructional teams are meeting specific achievement goals.
2- To determine whether a teacher’s students are meeting achievement growth expectations.
3- To gather information to measure what professional development opportunities are needed for individual teachers, teaching coordinators and grade-level leaders.
4- To gather evidence for making contract renewal and tenure decisions.
5- To determine whether a teacher’s performance qualifies him or her for additional compensation or a reward.
6- To determine who would qualify to become a phase or school leader or a departmental head.

2.2 Some historical studies for using Value-Added Method

Aaronson, Barrow and Sander (2007), attempted to estimate the importance of teachers in Chicago public high schools. They found that teachers' effectiveness was positively related to students' mathematics achievement, particularly for lower ability students. Lockwood, Louis and McCaffrey (2002), investigated the performance of rank or percentile estimators used to rank teachers based on students' achievement. Their findings showed that the use of Value-Added method to determine teacher ranking is inherently flawed. McCaffrey, Lockwood, Koretz and Hamilton (2004) presented research on Value-Added Method as part of a systematic review and evaluation of leading value-added approaches. In addition they discussed the use of Value-Added Methods for measuring teacher effectiveness, reviewed recent applications, and presented important statistical and measurement issues that might affect the validity of Value-Added Model inferences.

Mendro, Jordan, Gomez, Anderson and Bembry (1998), investigated the application of multiple linear regression techniques in determining longitudinal teacher effectiveness in which they found teacher impacts were related to students gains on the Iowa Test of Basic Skills (ITBS) test.

Noell (2006), examined the effect of teacher preparation programs on Value-Added methods of student achievement in English, mathematics, science and social studies in grades 4-9. Prior student achievement was the strongest predictor of Value-Added outcomes, and within-program variation was too high to detect any effects of preparation.

Rivers-Sanders (1999), examined residual and cumulative teacher impacts on student learning in mathematics for grade 4-8 and found that all students benefited from highly effective teachers, with the lower achieving benefiting most.

Rivkin, Hanushek and Kain (2005), investigated the influence of schools and teachers on students' achievement gains in mathematics and reading in grades 3-7. Unobserved differences in teacher quality accounted for most of the difference in achievement. Observable teacher characteristics showed some small effects.

Tekwe, Carter, Ma, Aligina, Lucas, Roth, et al (2004), investigated the impact of differences between three statistical models for assessing school performance using Value-Added method and found that there are correlations between the models and also some discrepancies.

Wright (2004), compared several different statistical approaches to value-added modeling in order to demonstrate the benefits of using a more complex, multivariate longitudinal approach to calculating value-added models. They presented the results from using each of these models.

2.3 Previous research on the effects of Value-Added Method

Ellett and Teddlie (2003), indicated that the development of sophisticated data analysis techniques such as statistical mixed model theory and development and hierarchical linear modeling now make it possible to not only conduct meaningful, longitudinal investigations of school and classroom level effect on school outcomes, but also to study the effect of within school variation on school effectiveness.

Newton, et al (2010), mentioned that the Value-Added Method has attracted much attention in the research and policy community as a hoped-for means to isolate the effect of teachers upon students' learning, raising the teacher's effect upon the other factors from other powerful factors. Many hope that if Value-Added Method can isolate teacher impacts on students' learning, then various personnel decisions can be based on the teacher impact estimates. Newton et al (2010) study highlighted the following strengths as being;

1- matching students with teachers at the course level, which allowed them to investigate the stability of teacher ranking cross different courses.
2- matching students with teachers with whom they studied for the entire academic year. This avoided the potential problem of attributing a student’s learning to a teacher when the student had not been under the care of the teacher over the entire academic year.
3- studying teachers at the high school level, which illuminated several practical and conceptual issues that have important policy implications for using value added estimates to hold schools or teachers accountable for student performance. They suggested that it is not simple to measure precisely how many individual teachers contribute to student learning, given the other factors involved in the learning process, the current limitation of tests, methods and educational system.

Braun (2005), summarized the progress made so far. He suggested a Value-Added Method moves the discussion about teacher quality to be centered on increasing student learning as the primary goal of teaching. It also can enhance a quantitative component, as well as by forcing the issue for researchers to reexamine questions of fairness and proper test use, which are important areas to be taken into consideration. Furthermore, by utilizing methods related to individual student growth, Value-Added Methods provide a more defensible foundation for teacher evaluation than is offered by methods based on the promotion of students meeting a fixed standard of performance. Nevertheless, it is a known fact that there have already been a number of investigations of different Value-Added Methods in a variety of settings. They have begun to give us a clearer picture of the strengths and limitations of the various approaches.

Slater, Davies and Burgess (2012), found that teachers matter a great deal. Having one-standard deviation better teachers raises the test score by 27% of a standard deviation. Having a good teacher as opposed to a poor teacher makes a big difference. Raising average teacher quality does seem a promising direction for public policy.

Milanowski (2011), concluded that there is a need to develop a practical method to translate school leaders' visions of effective instruction into explicit competency models. The model can then become the foundation for a set of practice methods that include observational rubrics for performance evaluation and management, performance assessment that would be part of tenure and pay systems, and walk through tools for day-to-day performance management and for evaluating the embedding of instructional strategies.

2.4 Difficulties and challenges of using Value-Added Method

Little, Goe and Bell (2009), identified the different problems that are related to comparing teacher effectiveness with improving students' learning as measured by student gains on standardized achievement tests. They highlighted in their research that:

- teachers are not exclusively responsible for students learning. It is affected by different factors such as other teachers, peers, family, home environment, school resources, community support, leadership and school climate and personal motivation. All contribute to a holistic other view of student learning.
- Consensus should drive research, not measurement innovations. It is possible that the increase in data linking students' achievement to individual teachers and new statistical techniques to analyze these data are contributing to an emphasis on measuring teacher effectiveness using student achievement gains.
- Test scores are limited in the information they can provide. Information is not available for some non-tested subjects and certain student population. Furthermore, basing teacher effectiveness on

student achievement fails to account for other important student outcomes. This method does not provide any additional information on student learning growth beyond the data gleaned through standardized testing, such as information about teachers whom teach elementary school, special education or untested subjects.

- Learning is more than just looking at the average achievement gains. Improving students' attitudes, motivation and confidence also contribute to learning and not only gains in achievement.

Hammond et al (2012), listed the different factors that influence gain in student achievement other than the teacher impact. These include:

- school factors such as class sizes, curriculum materials, instructional time and availability of specialist teacher and resources of learning.
- home and community support or challenge.
- individual student needs and abilities, health and attendance.
- peer culture and achievement.
- prior teachers and schooling.
- the natural of specific tests used.

2.5 The importance of using this method in schools in Dubai

From his own experience as an inspector the researcher has found that many schools are using different methods of evaluating teachers' performance. Very few of them could use the Value-Added Method as a tool to measure teacher's effectiveness.

Teacher evaluation is one of the different tools that schools use to monitor their improvement and teachers' effectiveness and renewal of their yearly contracts. Student's questionnaires, classroom observations by school leadership and teachers' portfolios are all part of the evaluation methods of a teacher's effectiveness.

Education continued to be an important issue and of high concern for UAE, which lead to the planning of a future picture for the development of education in the next twenty years, basing outcomes on higher standards and concentrating on science and technology. The plan incorporates information technology, computer illiteracy, preserving social values and ethics, and promoting traditional values among youth.

Most of the private schools in Dubai depend mainly on teachers that are hired from different countries outside UAE. Dubai a population of at least two million with at least 150 private schools offering 13 different curricula, change to different professional backgrounds and different teaching experiences. There were changes. There are more than 250,000 students studying at these schools This presents a challenge to educational policy makers to ensure that in this multicultural environment, schools are delivering higher outcomes for students.

Teacher evaluation is one of the different tools that schools use to monitor their improvement and to evaluate teachers' effectiveness prior to renewing their annual contracts. Until 2007 this process was very limited in schools commitment to this process as a part of their improvement plans or teachers development. Policies and legislation were not issued to take this into consideration.

The Dubai government has set few generic policies to assist with teachers' evaluation in classrooms or outside it. The policies that were set were mainly related to overall school evaluation and as part of the quality assurance of the schools' outcomes.

His Highness Sheikh Mohammad Bin Rashid Al Maktoum the Prime Minister and the Vice President of the UAE and the ruler of Dubai set out in January 2014 a national agenda for UAE. By 2021 the country would be among the 15 highest performing countries in Trends in International Mathematics and Science Study (TIMSS) and among the 20 highest performing countries in

Program for International Student Assessment (PISA). This set a high expectation on schools and especially on teachers to raise their expectations of students and hence a need for rigorous teacher evaluation. Using the Value-Added Method would facilitate and accelerate teacher evaluation and also lead to a raise in achievement. Accumulating data over the years about students' attainment related to specific teachers would help schools to better evaluate teachers' effectiveness.

Chapter 3 The Present Research

3.1 Methodology

DeMitchell's and Gagnon (2012), defined Value-added method as an inclusive term for a collection of complex statistical techniques that calculate the value a teacher adds to the education of a student through the use of multiple years of a student's test score data. This definition was also mentioned in Betebenner (2004), Heistad (1999), McCaffrey, Lockwood, Koretz, and Hamilton (2004), Noell, Porter and Patt (2007) and Valli, Croninger and Walters (2007).

It means to separate out the numerous non-educational factors, such as family background, that can possibly impact on a student's achievement, thus isolating and measuring the effects of teachers and schools. Value-added calculations compare a teacher's contribution to student's achievement with those of other teachers, making value-added method calculations simple. Teacher value-added scores are most often hierarchical, thus allowing comparisons of teachers by student outcomes without relation to normative data that provides value, therefore removing context from the equation. Hence, value-added data will differentiate the most effective teachers from the least effective ones and identify those in the middle. Teacher impacts are a fixed construct that is independent of the context of teaching. The successful use of value-added method requires a high level of confidence in the attribution of achievement gain to specific teachers.

Value added measures are calculated at student level. A student's attainment at end of key stage 4 is chiefly affected by two key factors: student's attainment at the end of key stage 2 and the impact of their secondary school up to the end of key stage 4. A value added measure aims to estimate the effect of the secondary school on the student's attainment taking their key stage 2 results into account.

Value added methods are estimated for each individual student by comparing their key stage 4 results with all other students with similar key stage 2 scores. The difference between a student's actual key stage 4 performance and their estimated key stage 4 performance gives the student their value added score. Each individual student value added score is relative to the performance of other students. Naturally some students will progress more than others independently of which school they attend and a school's value added score would certainly be different if they had a different cohort of students. Confidence intervals are provided as a proxy for a range within which you can be 95% certain the true value added score lies. These intervals should be taken into account when making comparisons between schools, groups or national averages. Therefore, the school value added measures that are published in the performance tables are presented alongside the respective 95% confidence intervals. School value added measures estimate the effect of the secondary school on all of their student's attainment.

McCaffrey et al (2004), mentioned that the recent literature on Value-Added Method suggests that teacher impacts on student learning are large, accounting for a significant portion of the variability in growth, and that they persist for at least three to four years into the future. It is difficult to determine the size of teacher impacts. Nonetheless, it appears that the magnitude of some of the effects reported in these studies is overstated. To determine the true size of teacher impacts, several important statistical and psychometric issues must be addressed.

1- Impact of basic issues of statistical modeling.
   This can have a significant impact on estimates of teacher performance. When the number of students taught by a particular teacher is small, estimates of teacher impacts can be heavily influenced by the performance of only few students. This can be solved by the modeling approach using a method called ‘shrinks’ in which it estimates for individual teachers back towards the overall mean.
2- Impact of omitted variables and missing data.
   Thus in Value-Added Method, analysts rely on observational, rather than experimental, data. Reliance on such data can lead to inaccuracy in estimates of teacher impacts due to (1) differences between schools or classrooms that are not fully controlled in the analysis (such differences ‘’confound’’ the results) such as students attending more than one school or even the school composition of different students and (2) gaps in the data collected in school.
3- Effect of using achievement tests as an outcome.
   Thus value-Added Method uses measures of students’ achievement to define and estimate teacher impacts, but these are limited in several ways. For example changes in the timing of the test, the weight given to alternative topics, or the methods used to create scores from students’ responses could affect conclusions about the relative achievement or growth in achievement across classes of students.
4- Effect of sampling error.
   Thus this is another source of error in Value-Added Method estimates. These are the estimates of teacher factors that have larger sampling errors than estimates of school effects because of the smaller numbers of students used in the estimation of the individual teacher impact.

Newton et al (2010) asserted that by using the previous different statistical controls, a teacher’s effectiveness is then measured by the average difference between end of year scores and IGCSE scores for all students assigned to that teacher. This method of teacher effectiveness has the advantage of transparency and is conceptually similar to estimates of the teacher fixed effects in more sophisticated Value-Added Method regression models. They recognize that stronger statistical controls would be possible using two or more prior years of student data for each teacher. However this has different practical limitations.

Many studies have been conducted to compare the different methodological approaches. In this research two methods are used to measure teacher's effectiveness, one is through the school evaluation of the teacher's performance and the other is through the gain in student's results from year 10 and year 11 in selected subjects.

Dossett and Munoz (2003) explained that multilevel analysis is just one of many methods for understanding data and it may not always work. Multilevel analysis is useful if the data is constructed similarly to the multilevel model. One variable that has shown to be important in previous research on school performance is prior academic achievement. Chubb & Moe (1990) and Smith & Meier (1995) found in many studies that performance on achievement tests is highly correlated with previous levels of academic achievement. Smith and Meier (1995) found that school systems doing well in the past continue to perform well. The study supported the hypothesis that previous academic performance is a strong predictor of future performance.

Roeder (1999) and (2000) studied the performance of schools in relationship to selected academic and social variables. A final set of factors are teacher-related variables, although common sense and schooling experiences suggest that teachers and teaching make a difference to students' achievement.

3.2 Research Instrument

The research instruments developed will measure the difference between students' results and achievement at the end of the year and their results in the international exams IGCSE for the same subject. They will check whether there is a positive difference which will mean that there is a teacher impact. These will then compare teacher impact with the school evaluation of that teacher for the same section and the same subject.

1- A table of male and female students' results and achievements in year 10 and year 11 for the end of year and IGCSE for biology, maths and physics is presented in Appendix 1.
2- A table of school evaluation of the different teachers for the previous year group and teaching the previous subjects is presented in Appendix 2.

1-Students' achievements and results:

A quantitative research is conducted using students' data and scores. The scores are analysed and interpreted in a process of a number of structured phases. The analysis used quasi-experimental research to measure if there is a teacher impact on the sample of the study. The other part is a qualitative method in which the school has provided the researcher with a list of evaluations of the individual teachers' performances.

The researcher has conducted the following different tests

a- Independent t-test.
b- ANOVA Test
c- Paired t-test

a- Independent t-test:

This test is used to compare results of the performance of the groups when making comparisons between the two groups. Using this test in an Excel program will provide useful information that can be easily interpreted. This test is used to show if there is a significant difference in the mean of these groups. It is used to determine whether means are significantly different in the two groups selected. This is the first test to be conducted on the results to help to interpret the data and describe the findings.

b- ANOVA Test:

ANOVA test (Analysis of Variance Test) is used to perform more than one comparison using the same set of data. It is flexible and easy to use to examine the differences between the set of scores for the end of year and the set of scores for the IGCSE for each year, subject, section and gender to get the differences among the variables in the mean and the standard deviation. ANOVA test is used to test the effect of teaching on the students' overall gain in their section scores. This test discovers the interactions between the independent variables that can measure the amount of variation within the scores in the multiple comparisons conducted.

c- Paired t-test:

The test is used to find paired sample correlations between variables to measure if there is significant decline or increase in the score of the mean or the standard deviation. The researcher will investigate the effect of size of the level of consistency. The scores will have to reflect the strength of correlation between the scores and hence make a generalization as to the findings and conclusions of the study.

2-School evaluation of the teachers;

The researcher obtained a table of the school evaluation grades for each teacher of the specified subject for each year group and for each section. This grade was arrived at through a classroom observation visit by a senior leader who recorded his/her judgments on the teacher's performance. However, the quality of the judgments depends on the observer's experience and on the evaluation criteria. The evaluation grade consists of four judgments levels 'Unsatisfactory', 'Acceptable', 'Good' and 'Outstanding'. The researcher was not given access to the evaluation form, nor the criteria used by observers to grade teachers. However the school did provide a list of evaluation grades for each subject by section and by year.

3.3 Sample:

The study sample consisted of 174 students (79 female and 95 male) from Year 10, who were studying two different subjects- biology and maths. Students were in 8 different sections as illustrated in Table 1. The other groups are made up of 192 students (75 female and 117 male) from Year 11. They were studying three subjects - biology, maths and physics. Students were divided into 8 different sections in one UK curriculum private school in Dubai, United Arab Emirates as illustrated in Table 2.

Year 10	Female	Male	Total
Biology Section 1	11	8	19
Biology Section 2	8	9	17
Sub-Total Biology	**19**	**17**	**36**
Math Section 1	10	13	23
Math Section 2	10	17	27
Math Section 3	11	14	25
Math Section 4	8	15	23
Math Section 5	10	9	19
Math Section 6	11	10	21
Sub-Total Math	**60**	**78**	**138**
Total Year 10	**79**	**95**	**174**

Table 1 Sample of students in Year 10 in different subjects and sections

Year 11	Female	Male	Total
Biology Section 1	10	10	20
Sub-Total Biology	**10**	**10**	**20**
Math Section 1	10	14	24
Math Section 2	7	15	22
Math Section 3	13	22	35
Sub-Total Math	**30**	**51**	**81**
Physics Section 1	9	15	24
Physics Section 2	6	13	19
Physics Section 3	12	11	23
Physics Section 4	8	17	25
Sub-Total Physics	**35**	**56**	**91**
Total Year 11	**75**	**117**	**192**

Table 2 Sample of students in Year 11 in different subjects and sections

The other sample is of 16 teachers from the same school, teaching different subjects to students in different sections. This is illustrated in Table 3.

Teachers	Female	Male
Y10 Biology Section 1	1	-
Y10 Biology Section 2	1	-
Y10 Math Section 1	1	-
Y10 Math Section 2	-	1
Y10 Math Section 3	1	-
Y10 Math Section 4	-	1
Y10 Math Section 5	1	-
Y10 Math Section 6	1	-
Sub-Total Y10	**6**	**2**
Y11 Biology Section 1	1	-
Y11 Math Section 1	-	1
Y11 Math Section 2	-	1
Y11 Math Section 3	1	-
Y11 Physics Section 1		1
Y11 Physics Section 2	-	1
Y11 Physics Section 3	-	1
Y11 Physics Section 4	1	-
Sub-Total Y11	**3**	**5**
Total	**9**	**7**

Table 3 Sample of teachers for Year 10 and Year 11 for the different subjects and sections

The sample of students and teachers was selected at random. The school supplied student numbers for each section in year 10 and 11 together with their gender and end of year and IGCSE scores for the academic year 2013-2014. They also provided evaluations and gender information for the teachers' sample group.

3.4 Ethics

In such a type of study ethics is an important issue. After gathering all information and analyzing it, some ethical issues should be taken into consideration. Teachers were notified by the school that their evaluations and gender with their names would be included with this study and agreed to participate. Great care should be taken to ensure that all of the sample names and any other identifying information would not be assigned to any person.

3.5 Study Design and Data Collection Method

Fraenkel and Wallen identified Quantitative research as:

> "A research in which the researcher attempts to analyze a situation through a designed and controlled set of data collection and analysis" (2009, p10).

Quantitative research is based on the desire to understand the why of situations and on the belief that social situations can be measured and clearly expressed in numbers. This allows researchers to arrange and group variables used to give an explanation of the action and to make comparisons and to measure them.

Qualitative research on the other hand gives an answer to the how questions, and explaining a situation. It raises questions on previously unknown answers. It takes into account an individual's opinions and in addition considers immeasurable experiences on a specific subject. Such data is collected through interviews, questionnaires and observations. These data use words that are similar to measurable units. Such data are set into a clear structure for correct interpretation.

In order to answer the question of teacher effectiveness and students' improved achievement, the researcher intended to use a mixed methods case study that uses both qualitative and quantitative methods. Hence, the researcher will use a quantitative data analysis of students' end of year results and the IGCSE results in different subjects for years 10 and 11, as well as a judgment on the qualitative observation of the participating teachers in the study conducted by the senior leader of the school.

Creswell and Plano Clark define the mixed methods research design approach as:

> "Mixed methods research is a research design with philosophical assumptions as well as methods of inquiry. As a methodology, it involves philosophical assumptions that guide the direction of the collection and analysis of data and the mixture of qualitative and quantitative approaches in many phases in the research process. As a method, it focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone (2007, p.5)".

In this way, the two methods used will triangulate the findings in this case study so it will be more in-depth and reliable. Data collected in this study are of different types. This will provide opportunities for further explanation of the two methods. In addition, the researcher will be able to measure the variable used and to take the study context into consideration. This combination of data will give the researcher the opportunity to analyze and discover the relationship between the variables used and hence to measure teachers' effectiveness.

The process of trying to obtain data from most of the schools in Dubai was fraught with difficulty. The researcher gathered the names of different UK private schools to target, i.e. schools with large numbers of students in years 10 and 11. Most private schools in Dubai are reluctant to cooperate with Master or Doctoral students, but the study supervisor identified a possible principal who could be approached. However, the principal was unwilling to participate due to the quantity of information required. Another school was identified some two weeks later using the researcher's work contacts and the principal agreed to participate, but was only prepared to give specific information for students from years 10 and 11 for 2013-2014. This appears in appendix 1. In addition, a list of teachers for the different sections was provided together with details of gender and evaluation grades. This information is in Appendix 2. It was made clear that no other information would be provided due to confidentiality constraints. It was agreed by the supervisor that the study could proceed with the limited amount of information.

3.6 Case Studies

Fraenkel and Wallen define a case study as a research approach in which an individual or group is “studied extensively and varied data are collected and used to formulate interpretations applicable to the specific case or to provide useful generalizations” (2009, p.13). The primary purpose of such a study is to allow the researcher to develop a theory and make insights based upon the resulting data and to take into account several different factors including the school, teacher and students contexts occurring with the selected group of participants. Case studies are usually holistic in nature as they accumulate numerous details as they focus on the individual or group being studied in a real-life context. They can be based on either quantitative or qualitative research or both, and collect data through numerous methods during the same study. Advantages of using case studies in research are that by intensely focusing on a specific individual or group over a set period of time, a researcher can learn large amounts of information, including details and information otherwise not expected. This, in turn, can allow the researcher to better understand all the factors involved with and affecting the participants, including those factors that may have been previously unexpected by the researcher.

3.7 Reliability, Validity and Triangulation

Both reliability and validity are of utmost importance and are the foundations to any study being considered credible. According to Fraenkel and Wallen (2009), reliability refers to the “consistency of the scores obtained – how consistent they are for each individual from one administration of an instrument to another, and from one set of items to another” (2009, p. 147). In other words, for results to be considered reliable, the different instruments used should consistently produce the same results over time. A participating individual’s results should also be able to be confirmed using the test-re-test method at a later time (Golafshani 2003). This is described by Joppe (n.d.) who defined reliability as:

> ''The extent to which results are consistent over time and an accurate representation of the total population under study is referred to as reliability. In other words, if the results of a study can be reproduced under a similar methodology, then the research instrument is considered to be reliable'' (n.d., p.1).

Validity refers to the ability of a test to measure what it was designed and intended to measure. It can also be described as the degree to which appropriate inferences can be made based on the results produced by an instrument. Flick (2007) and Fraenkel and Wallen (2009). Validity depends on more than just the research instrument alone. It also depends on the instrumentation process as well as the characteristics of the group used as the participant sample. Triangulation refers to the use of different research methods, methods of data collection, and/or the use of different types of data to study the same research question and then converge on a single understanding and interpretation Flick (2007) and Fraenkel and Wallen (2009).Triangulation in mixed methods research is successfully achieved when both the quantitative and qualitative data are integrated and both validate and confirm each other.

Chapter 4 Findings and Discussion

The purpose of this chapter is to report on the research findings from the series of tests conducted during the study. The data from these tests have been analysed in order to determine the statistical significance of the scores. They will also serve to measure whether there is a gain in students' scores between the end of year results and the IGCSE scores and what, if any, impact the teacher has had on these scores. This gain is a factor known as Value-Added. The research answers the three questions. All the information provided by the school for the study is described in Chapter 3. As mentioned in Chapter 3 a quasi-experimental approach is conducted to find answers to the research questions using different tests;

- Independent t-test.
- One way ANOVA Test
- Paired t-test.

The process of analysis includes applying the previous tests to perform the following subsequent tests to each set of scores for Year 10 and Year 11 for each subject and for each section and finally compare the result with the evaluation judgment for each teacher.

1- The independent t-test is conducted first to the different data to check if there is a difference in the mean for all students in this set,
2- Then ANOVA test is applied to the same set of data to check if there is a significant difference in the standard deviation for each set's end of year scores or IGCSE scores and if this difference applies to the IGCSE scores.
3- Thirdly the paired t-test is used to find the correlation between the two variables of the end of the year and IGCSE results and whether there is a significant decline or increase in the results mean or standard deviation thereby confirming the effect of teacher on this group.
4- Finally this section analysis which has a teacher impact will be compared with the list of teachers' evaluations provided by the school.

This will support the hypothesis that a gain in the Value-Added model is proving a teacher's effect on that section.

5- The same processes are applied to apply to female and male students in each section.

The research questions are;

1- Could the positive or negative differences of students' results in Year 10 and 11 in maths, physics and biology in IGCSE exams when compared with end of year results be related to effective teaching? Hence could it be a measure for teacher effectiveness?
2- Is there a gender implication, for example is there a difference between the overall performance of male and female students when their teacher is a male or a female?
3- Is there is a difference between value added method findings and the evaluation of teachers conducted by the senior leadership team of the same teacher for the same year group and the same subject?

4.1 Summary of data analysis:

<u>In year 10;</u>

- In the biology sections, the overall results showed that there is no significant improvement in students' results in IGCSE. No teacher's impact was found in two biology sections, although the teachers of these sections were evaluated 'Good' by the school.
- In the maths sections, overall results showed that there is a significant improvement in students' IGCSE results compared to those at the end of the year. Only two sections (1) and (6) out of six showed this improvement and those classes were taught by two different 'teachers evaluated as 'good' by the school.
- The results show that the different biology sections there is no to teacher effect on either male or female students. However in maths the teacher effect results in both male and female gains. The only female effect

raised in student’s results section (6) which is taught by a ‘Good’ female teacher.

<u>In year 11;</u>

- In the biology section, there is a significant improvement in students’ IGCSE scores compared with EOY results. This indicates there is a teacher impact on the section where the teacher was graded ‘Good’.
- In the maths sections, overall results showed that there is a significant improvement in IGCSE results compared to EOY. The improvement was noted in all three sections. These classes were taught by three different who had been graded Acceptable.
- In the physics sections, overall results showed that there is no significant difference in either students’ IGCSE or EOY results This would indicate that these teachers have no impact on students’ achievement. In these classes, there were two different teachers, one graded ‘Good’, the other ‘Acceptable’
- The results show that there is a teacher effect in the only biology section. There was evidence that the teacher, a female teacher graded ‘Good’, had an impact on the achievement of male students but not on female students. In maths section 1- we have seen a change in the results for both female and male groups taught by an ‘Acceptable’ male teacher. In sections 2 and 3, taught by a male teacher and a female teacher respectively both of whom had been graded ‘Acceptable’, the teacher effect on the male students’ scores was evident. In the four sections in physics there was no evidence of teacher impact on either male or female groups. These classes were taught by male and female teachers graded ‘Good’ or ‘Acceptable’.

Overall, results agree with research question one and to a certain extent question number three. The research findings did not provide an answer for question two which would suggest the need for more research into a link between teacher gender and the impact on students' overall results.

4.2 Overall Analysis;

In this section the researcher will answer questions 1 and 2 above.

4.2.1 Year 10 analysis;

Table 4.1 illustrates the mean of 174 students (79 females and 95 males) for the end of year (EOY) scores and IGCSE scores for the two subjects -biology and maths- after performing independent t-test and ANOVA test to get the standard deviations (Std. Dev.) for the end of year and IGCSE results with the same analysis performed for female and male students;

Subject	Number of Students	Mean EOY	Std. Dev. EOY	Mean IGCSE	Std. Dev. IGCSE
Biology	36	5.666	1.01	5.638	1.150
Female	19	5.631	1.065	5.526	1.172
Male	17	5.706	1.014	5.765	1.195
Math	138	5.811	0.970	5.586	1.092
Female	60	5.833	0.866	5.550	1.060
Male	78	5.795	1.011	5.615	1.119

Table 4.1 Year10 analysis of biology and maths including female and males for t-test and ANOVA test

The analysis shows that:

- The standard deviation for the IGCSE scores for biology and maths is greater than that for end of year scores for the two subjects.
- The standard deviation for the IGCSE scores for biology and maths is greater than that of end of year scores for the two groups of females and males.

Table 4.2 illustrates the Paired t-test analysis of 174 students (female and male) in which t-statistical (t-stat) value is compared with the t-calculated (t-calc.) value of two tail with a confidence interval of 95% and one tail value with the same confidence interval.

Subject	Number of Students	t-stat	t-calc. one tail	t-calc. two tail	Teacher impact
Biology	36	0.171	2.437	2.030	No teacher impact
Female	19	0.437	1.734	2.100	No teacher impact
Male	17	-0.269	2.76	2.119	No teacher impact
Math	138	2.892	1.656	1.977	Yes teacher impact
Female	60	2.210	1.671	2.000	Yes teacher impact
Male	78	1.867	1.656	1.991	Yes teacher impact

Table 4.2 Year10 analysis of biology and maths with female and males for paired t-test and teacher impact.

This analysis shows that:

- The t-stat for the biology is less than the t-calc. one tail and two tail values.
- The t-stat for the maths is greater than the t-calc. one tail and two tail values.

The results show that:

- Maths students in table 4.1 will gain a change, which is an improvement. Therefore there should be a teacher impact in one or more sections of Year 10 maths.
- The t-statistical in biology and in female and male groups is less than t-calculated for the one tail and two tail values, therefore no improvement, hence and no teacher impact.
- The t-statistical in maths and in female and male groups is greater than t-calculated for the one tail and two tail values, therefore there is an improvement, hence and there is a teacher impact.

4.2.1.1 Year 10 analysis for biology sections;

Table 4.3 illustrates the means of the 36 students (19 female and 17 male) in the two sections for end of year scores and IGCSE scores for biology after performing independent t-test and ANOVA test to arrive at the standard deviations for the end of year and IGCSE results.

Biology	**Number of Students**	**Mean EOY**	**Std. Dev. EOY**	**Mean IGCSE**	**Std. Dev. IGCSE**
Section 1	19	5.578	0.837	5.473	1.218
Female	11	5.545	0.934	5.364	1.206
Male	8	5.625	0.744	5.625	1.302
Section 2	17	5.764	1.251	5.823	1.074
Female	8	5.75	1.165	5.75	1.282
Male	9	5.777	1.054	5.889	1.302

Table 4.3 Analysis of biology sections including female and male groups for t-test and ANOVA test

The analysis shows that:

- The standard deviation for the IGCSE scores for biology section one is greater than EOY.
- In biology section two it was the reverse.
- For the female and male groups, the standard deviation for IGCSE scores in the two groups in the two sections is greater than EOY values.

Table 4.4 illustrates the Paired t-test analysis of the 36 students (19 female and 17 male) in the two sections in which t-statistical value is compared with the t-calculated value of two tail with a confidence interval of 95% and the t-calculated value of one tail with the same confidence interval.

Biology	Number of Students	t-stat	t-calc. one tail	t-calc. two tail	Teacher impact
Section 1	19	0.566	1.689	2.100	No teacher impact
Female	11	0.690	1.812	2.228	No teacher impact
Male	8	0	1.894	2.364	No teacher impact
Section 2	17	0.211	1.745	2.110	No teacher impact
Female	8	0	1.895	2.365	No teacher impact
Male	9	0.316	1.859	2.306	No teacher impact

Table 4.4 Analysis of biology sections including female and male groups using paired t-test and showing teacher impact

The analysis shows that:

- The t-statistical for the two biology sections is less than the t-calculated one tail and two tail values.
- The t-statistical for the two biology sections for female and male groups is less than the t-calculated one tail and two tail values.

The results show that:

- Biology students in table 4.3 support section one to have an effect. However the two sections in table 4.4 will not gain a change therefore there is no teacher impact.
- The female and male groups in both sections will not gain a change therefore there should be no teacher impact for all groups of females and males in the two sections.

4.2.1.2 Year 10 analysis for maths sections.

Table 4.5 illustrates the means of 138 students (60 female and 78 male) distributed into six sections. It also shows the end of year scores and IGCSE scores for maths after performing independent t-test and ANOVA test to arrive at the standard deviations for the end of year and IGCSE in addition to that of the standard deviation of females and males groups.

Math	Number of Students	Mean EOY	Std. Dev. EOY	Mean IGCSE	Std. Dev. IGCSE
Section 1	23	5.826	1.029	4.478	1.081
Female	10	6.1	0.876	5.6	1.075
Male	13	5.615	1.121	5.385	1.121
Section 2	27	5.962	0.979	5.555	0.891
Female	10	6.2	0.789	5.5	0.707
Male	17	5.823	1.074	5.588	1.004
Section 3	21	5.714	0.956	5.761	0.889
Female	11	5.727	1.009	5.909	0.94
Male	10	5.7	0.949	5.6	0.843
Section 4	25	5.680	1.029	5.600	1.220
Female	11	6.727	0.904	5.727	1.348
Male	14	5.643	1.150	5.5	1.160
Section 5	23	5.739	0.810	5.478	1.377
Female	8	5.5	0.926	5.25	1.389
Male	15	5.867	0.743	5.6	1.404
Section 6	19	5.947	0.911	5.684	1.108
Female	10	5.7	0.675	5.2	0.918
Male	9	6.222	1.093	6.222	1.093

Table 4.5 Analysis of maths sections including female and male for t-test and ANOVA test

The analysis shows that:

- The standard deviation for the IGCSE scores for maths is greater than the EOY in sections one, four, five and six.
- In sections two and three it was the reverse.
- For the female and male groups in section one, the standard deviation of females in IGCSE results is greater than EOY.
- In sections four and five for female and males groups, the standard deviations for the IGCSE are greater than EOY.
- The standard deviation for the female group in section six is greater than EOY.
- There were no female and male effects in sections two and three.

Table 4.6 illustrates the Paired t-test analysis for the 138 students (60 female and 78 male) in each section in which t-statistical value is compared with the t-calculated value of two tail with a confidence interval of 95% and the t-calculated value of one tail with the same confidence interval.

Math	Number of Students	t-stat	t-calc. one tail	t-calc. two tail	Teacher impact
Section 1	23	1.885	1.717	2.073	Yes teacher impact
Female	10	1.627	1.833	2.262	No teacher impact
Male	13	1	1.782	2.178	No teacher impact
Section 2	27	2.096	1.705	2.055	No teacher impact
Female	10	2.089	1.833	2.262	No teacher impact
Male	17	1	1.745	2.119	No teacher impact
Section 3	21	-0.27	1.724	2.080	No teacher impact
Female	11	-0.614	1.812	2.228	No teacher impact
Male	10	0.557	1.833	2.262	No teacher impact
Section 4	25	0.401	1.710	2.063	No teacher impact
Female	11	0	1.812	2.228	No teacher impact
Male	14	0.563	1.771	2.160	No teacher impact
Section 5	23	1.238	1.717	2.073	No teacher impact
Female	8	0.607	1.894	2.365	No teacher impact
Male	15	1.074	1.761	2.145	No teacher impact
Section 6	19	1.755	1.734	2.100	Yes teacher impact
Female	10	3	1.833	2.262	Yes teacher impact
Male	9	0	1.859	2.306	No teacher impact

Table 4.6 Analysis of Math sections including females and males groups for paired t-test and teacher impact

The analysis shows that:

- The t-statistical value for the three maths sections one, two and six is greater than the t-calculated one tail values.
- The t-statistical value for the three maths sections three, four and five is less than the t-calculated one tail values.
- The t-statistical value for female and males groups from the different sections is less than the t-calculated one tail and two tail values for all sections except for female group in section six where the t-statistical is greater than t-calculated one and two tail values.

The results show that:

- Maths students in table 4.5 supports that sections one and six gain change therefore there is an improvement in student's results meaning there is a teacher impact.
- For the female and males groups from the different sections the gain in student's results is for the female group values in section six therefore there will be a teacher impact in this section only.

4.2.2 Year 11 analysis;

Table 4.7 illustrates the mean of 192 students (75 females and 117 males) for the end of year scores and IGCSE scores for the three subjects- biology, maths and physics after performing independent t-test and ANOVA test to get the standard deviations for the end of year and IGCSE results. The same analysis was performed for female and male groups for each subject;

Subject	Number of Students	Mean EOY	Std. Dev. EOY	Mean IGCSE	Std. Dev. IGCSE
Biology	20	5.65	0.875	5.25	1.251
Female	10	5.4	0.843	5.3	1.18
Male	10	5.9	0.875	5.2	1.135
Math	81	5.987	0.844	4.604	1.594
Female	30	6.1	0.843	5.06	1.418
Male	51	5.921	0.875	4.33	1.135
Physics	91	5.824	0.972	5.703	2.019
Female	35	5.828	0.785	5.886	1.658
Male	56	5.821	2.222	5.589	1.080

Table 4.7 Year11 analysis of biology, math and physics including female and males for t-test and ANOVA test

The analysis shows that:

- The standard deviation for the IGCSE scores for three subjects is higher than that of the end of year scores.

- The same is true for the female and male in the three subjects where the standard deviation of the IGCSE is greater than the EOY.
- The standard deviation for the IGCSE score is less than EOY for only the male group in physics.

Table 4.8 illustrates the Paired t-test analysis of 192 students (75 female and 117 male) in which t-statistical value is compared with the t-calculated value of two tail with a confidence interval of 95% and one tail value with the same confidence interval.

Subject	Number of Students	t-stat	t-calc. one tail	t-calc. two tail	Teacher impact
Biology	20	2.179	1.729	2.093	Yes teacher impact
Female	10	0.361	1.833	2.262	No teacher impact
Male	10	2.279	1.833	2.262	Yes teacher impact
Math	81	9.114	1.664	1.990	Yes teacher impact
Female	30	6.998	1.699	2.045	Yes teacher impact
Male	51	7.192	1.675	2.008	Yes teacher impact
Physics	91	0.568	1.661	1.986	No teacher impact
Female	35	0.191	1.690	2.032	No teacher impact
Male	56	0.795	1.673	2.004	No teacher impact

Table 4.8 Analysis of Year 11 scores for biology, math and physics with female and male groups for paired t-test and teacher impact.

The analysis shows that:

- The t-statistical for biology and maths is greater than the t-calculated one tail and two tail values.
- The t-statistical for physics is less than the t-calculated one tail and two tail values.
- The t-statistical for female and male groups is greater than t-calculated for one tail or/and two tail values in the male group in biology and both female and male groups in maths.

- The t-statistical for female and male groups is less than t-calculated for one tail and two tail values in female group in biology and both female and male groups in physics.

The results show that:

- The teacher effect was evident in biology and maths students in Table 4.7. Nevertheless the two subjects in table 4.8 show an improvement, therefore there has been a teacher impact.
- The male group in biology and the two female and male groups in maths will gain a change hence there should be a teacher impact for these groups.

4.2.2.1 Year 11 analysis for biology section;

Table 4.9 illustrates the means of the 20 students (10 female and 10 male) in the biology section for end of year and IGCSE scores after performing independent t-test and ANOVA test to get these values for female and male groups;

Biology	**Number of Students**	**Mean EOY**	**Std. Dev. EOY**	**Mean IGCSE**	**Std. Dev. IGCSE**
Section 1	20	5.65	0.875	2.25	1.251
Female	10	5.4	0.843	5.3	1.418
Male	10	5.9	0.875	5.2	1.135

Table 4.9 Analysis of biology section including female and male groups for t-test and ANOVA test

The analysis shows that:

- The standard deviation for the IGCSE scores for biology section is greater than EOY.
- The standard deviation for IGCSE scores for the female and male groups is greater than EOY values.

Table 4.10 illustrates the Paired t-test analysis of the 20 students (10 female and 10 male) in biology section in which t-statistical value is compared with the t-calculated value of two tail with a confidence interval of 95% and the t-calculated value of one tail with the same confidence interval.

Biology	Number of Students	t-stat	t-calc. one tail	t-calc. two tail	Teacher impact
Section 1	20	2.179	1.729	2.093	Yes teacher impact
Female	10	0.361	1.833	2.262	No teacher impact
Male	10	2.279	1.833	2.262	Yes teacher impact

Table 4.10 Analysis of biology section including female and male groups using paired t-test and teacher impact

The analysis shows that:

- The t-statistical for the biology section is greater than the t-calculated one tail and two tail values.
- The t-statistical for the male group is greater than the t-calculated one tail and two tail values.
- The t-statistical for the female group is less than the t-calculated one tail and two tail values.

The results show that:

- The teacher effect was noticeable in Table 4.9. Therefore students in this section will show an improvement and so there should be a teacher impact.
- In this section it showed that there will be a raise in student's results for the male group but not the female group, demonstrating teacher impact on the male group only.

4.2.2.2 Year 11 analysis for maths sections;

Table 4.11 illustrates the means of 81 students (30 female and 51 male) distributed into three sections for the end of year and IGCSE scores for maths after performing independent t-test and ANOVA test to get the standard deviations for the end of year and IGCSE scores in addition to the standard deviation for female and male groups.

Math	**Number of Students**	**Mean EOY**	**Std. Dev. EOY**	**Mean IGCSE**	**Std. Dev. IGCSE**
Section 1	24	5.875	0.679	4.666	1.761
Female	10	5.8	0.632	5.1	1.287
Male	14	5.928	0.730	5.357	2.023
Section 2	22	6.136	0.888	4.590	1.763
Female	7	6	1.291	5	1
Male	15	6.2	0.676	4.4	2.028
Section 3	35	5.971	0.923	4.571	1.399
Female	13	6.385	0.869	5.076	0.862
Male	22	5.727	0.882	4.273	1.579

Table 4.11 Analysis of maths sections including female and male for t-test and ANOVA test

The analysis shows that:

- The standard deviation for the IGCSE scores for maths is greater than the EOY scores in all sections.
- The standard deviation for the IGCSE scores for female and male groups is greater than EOY scores in section one female and male groups, and male groups in sections two and three.

Table 4.12 illustrates the Paired t-test analysis for the 81 students (30 female and 51 male) in each section where t-statistical value is compared with the t-calculated value of two tail with a confidence interval of 95% and the t-calculated value of one tail with the same confidence interval.

Math	Number of Students	t-stat	t-calc. one tail	t-calc. two tail	Teacher impact
Section 1	24	3.938	1.713	2.068	Yes teacher impact
Female	10	2.089	1.833	2.262	Yes teacher impact
Male	14	3.465	1.771	2.160	Yes teacher impact
Section 2	22	4.543	1.720	2.079	Yes teacher impact
Female	7	4.582	1.943	2.447	Yes teacher impact
Male	15	3.749	1.761	2.145	Yes teacher impact
Section 3	35	7.416	1.690	2.032	Yes teacher impact
Female	13	7.749	1.782	2.178	Yes teacher impact
Male	22	5.108	1.720	2.079	Yes teacher impact

Table 4.12 Analysis of maths sections including females and males groups for paired t-test and teacher impact

The analysis shows that:

- The t-statistical value for the three maths sections is greater than the t-calculated one tail and two tail values.
- The t-statistical value for female and male groups in all three sections is greater than t-calculated for one and two tail values.

The results show that:

- Maths students in all sections show an improvement in their scores, demonstrating teacher impact..
- Female and male groups in the different sections made gains in their results, indicating a teacher impact on different groups in all sections.

4.2.2.3 Year 11 analysis for physics sections;

Table 4.13 illustrates the means of 91 students (35 female and 56 male) distributed into four sections for the end of year and IGCSE scores for physics after performing independent t-test and ANOVA test to arrive at the standard deviations for the end of year and IGCSE scores in addition to the standard deviation of female and male groups.

Physics	Number of Students	Mean EOY	Std. Dev. EOY	Mean IGCSE	Std. Dev. IGCSE
Section 1	24	5.958	0.690	5.833	1.926
Female	9	5.778	2.179	5.333	0.667
Male	15	6.067	1.767	6.133	0.703
Section 2	19	5.736	1.593	5.157	2.794
Female	6	5.5	1.049	6.166	0.753
Male	13	5.846	1.818	4.692	3.275
Section 3	23	5.782	0.735	5.956	1.941
Female	12	5.916	0.514	6.333	0.778
Male	11	5.636	0.924	5.545	1.968
Section 4	25	5.8	0.816	5.76	1.877
Female	8	6	1.069	5.625	2.386
Male	17	5.706	0.686	5.823	1.667

Table 4.13 Analysis of physics sections including female and male for t-test and ANOVA test

The analysis shows that:

- The standard deviation for the IGCSE scores for physics is greater than the EOY in all sections.
- In section one the standard deviation of IGCSE scores for female and male groups is less than EOY values.
- In sections two and three, the standard deviation for the male group for the IGCSE is greater than EOY value.
- Finally in section four the standard deviation for the female and male group for IGCSE is greater than EOY value.

Table 4.14 illustrates the Paired t-test analysis for the 91 students (35 female and 56 male) in each section in which t-statistical value is compared with the t-calculated value of two tail with a confidence interval of 95% and the t-calculated value of one tail with the same confidence interval.

Physics	Number of Students	t-stat	t-calc. one tail	t-calc. two tail	Teacher impact
Section 1	24	0.289	1.713	2.068	No teacher impact
Female	9	0.609	1.859	2.306	No teacher impact
Male	15	-0.121	1.761	2.144	No teacher impact
Section 2	19	1.027	1.734	2.100	No teacher impact
Female	6	-3.162	2.015	2.570	No teacher impact
Male	13	1.489	1.782	2.179	No teacher impact
Section 3	23	-0.491	1.717	2.073	No teacher impact
Female	12	-2.159	1.796	2.200	No teacher impact
Male	11	0.126	1.812	2.228	No teacher impact
Section 4	25	0.103	1.710	2.063	No teacher impact
Female	8	0.382	1.894	2.365	No teacher impact
Male	17	-0.334	1.746	2.119	No teacher impact

Table 4.14 Analysis of physics sections including females and males groups for paired t-test and teacher impact

The analysis shows that:

- The t-statistical values for the four physics sections is less than the t-calculated one tail and two tail values
- The t-statistical values for all groups of female and males for the different sections is less than the t-calculated one and two tail values.

The results show that:

- The standard deviation for the IGCSE score for all students in the physics sections is less than EOY score meaning that. The t-statistical value is less than t-calculated one tail and two tail value. This shows that there is no improvement in student's t-statistical scores and therefore no teacher impact.
- The t-statistical value for male and female groups is less than t-calculated for one tail and two tail values. This indicates that there is no improvement and therefore there will be no teacher impact in any section or groups.

4.3 Comparisons between Value-Added Method results and school results

In this section the researcher will answer question 3 mentioned in the beginning of this chapter. The researcher will compare the results between the value added method findings and the evaluation of teachers in the same year group and the subject.

Table 4.3.1 shows a comparison between the value-added model results obtained previously and the teacher evaluations.

Subject's Sections	**Teacher's Gender**		**Value-added model results**	**Teacher's evaluation**
	Female	Male		
Y10 Biology Section 1	1	-	No teacher impact	Good
Y10 Biology Section 2	1	-	No teacher impact	Good
Y10 Math Section 1	1	-	Yes teacher impact	Good
Y10 Math Section 2	-	1	No teacher impact	Good
Y10 Math Section 3	1	-	No teacher impact	Acceptable
Y10 Math Section 4	1	-	No teacher impact	Good
Y10 Math Section 5	-	1	No teacher impact	Good
Y10 Math Section 6	1	-	Yes teacher impact	Good
Y11 Biology Section 1	1	-	Yes teacher impact	Good
Y11 Math Section 1	-	1	Yes teacher impact	Acceptable
Y11 Math Section 2	-	1	Yes teacher impact	Acceptable
Y11 Math Section 3	1	-	Yes teacher impact	Acceptable
Y11 Physics Section 1	-	1	No teacher impact	Acceptable
Y11 Physics Section 2	-	1	No teacher impact	Acceptable
Y11 Physics Section 3	-	1	No teacher impact	Acceptable
Y11 Physics Section 4	1	-	No teacher impact	Good

Table 4.3.1 shows the comparisons between Value-Added Method results with teacher's evaluations judgments from the school.

The results show that:

- Out of eight sections in year 10, two have shown teacher impact because of the value added results
- This correlates with the school's 'Good' evaluation of these teachers for these sections
- In year 11, however, four out of eight sections have produced similar results, but with one teacher graded 'Good' and the other three graded 'Acceptable'

Chapter 5 Conclusion and Recommendations

5.1 Conclusion;

The study is the first of its kind in Dubai and the UAE. It investigates whether the quality of a teacher's practice adds value to student achievement, and therefore whether value added can be used as a reliable measure of a teacher's effectiveness.

The Value-Added Method has been shown to be a good tool to measure teacher effectiveness and the impact that teaching has on learning. The statistical analysis using the quasi-experimental approach, and the series of tests applied to the data reveal - that there was either a positive change or no change in students' overall results. Therefore, the research is inconclusive. However, what the research did not take account of was the teachers' knowledge of the examination requirements and their ability to incorporate that into their teaching.

In addition, when comparing these findings with teacher evaluations, Slater, Davies and Burgess (2012), concluded that good or poor teachers can have an impact on students' test results. A good teacher with a one-standard deviation can raise a test score by 27% of a standard deviation. Having a good teacher or poor teacher makes a big difference.

Newton, Hammond, Haertel and Thomas (2010), concluded that it is not that teachers do not matter. Rather, their findings suggest that this simply cannot measure precisely how many individual teachers contribute to students' learning, given the other factors involved in the learning process and due to the current limitations of tests and methods. Other studies are needed to evaluate these issues further, and to develop strategies for taking into account the various factors that may influence the gains in students' achievement, so that the effect of students' learning can be properly understood.

The data required to carry out this study were complex and difficult to obtain, and not sufficiently extensive. Nevertheless, repeating or extending this research to other schools and other curricula in private and governmental schools in Dubai and UAE is of great value.

5.2 Limitations

The research as mentioned previously is to test whether an improvement in students' achievement results is because of teacher impact and to relate this outcome to an evaluation of the teacher's performance in the school. The findings revealed that there is an effect although it is minimal and hence the value-added model is one of the successful tools to measure teacher effectiveness.

There are limitations to all factors affected this research. All these should be taken into consideration in future research.

1- Data collected were not selected by the researcher but provided by the school.
2- Information related to students' nationalities, age and the length of time they have been studying in this school with the same teacher were not provided by the school.
3- Information relating to teachers' backgrounds and experiences was not provided.
4- Information on the end of year assessments used for comparison was not provided. There was no indication as to how the level or the marking criteria for these tests compared to the IGCSE.
5- No information was provided on the observation process, leaving the researcher questioning its reliability and validity. The criteria used for observing and grading teachers were not produced, nor was it clear whether teachers were aware of them. In addition the observations were

carried out by different senior leaders and did not appear to have been moderated.

6- The researcher used the quasi-experimental approach that provides statistical analysis focusing on the analysis of the results. However questions remains over the quality and accuracy of the observations carried out by senior leaders.
7- Lack of any research undertaken in this filed in Dubai or in the UAE that could help researchers to identify the problems and provide solutions.
8- Insufficient access to staff records on continuous professional development and the impact that this may have had on improving the quality of teaching.

5.3 Recommendations

This study could be of use to decision makers in determining how to evaluate a teacher's performance. While the study has its limitations, value-added model is considered to be one of the most important approaches with which to evaluate and measure a teacher's effectiveness.

Ko, Sammons and Bakkum (2013), suggested that some teachers may be better at teaching certain subject, or delivering certain topics or subject areas, or meeting certain teaching objectives. Teacher impacts may not be consistent but may fluctuate over the school year, across different phases of implementation of an educational policy, across different teaching periods and across lessons in which observation/assessment has taken place. There is a need to;

1- Adopt a broader definition of teacher effectiveness that emphasises the promotion of students' academic and other kinds of educational outcomes.

2- Understand that effective teaching is not automatically guaranteed through teacher training, professional development or long years of experience.
3- Recognize that teachers may vary in their effectiveness across years and in achieving different kinds of student outcomes and in teaching different groups of students or in different school contexts.
4- Disseminate and study relevant research and where available, inspection evidence on effective teaching practice and evaluate their applicability in different class and school contexts.
5- Identify and disseminate examples of successful practice from case studies of the work of effective teachers, effective departments and effective schools, through appropriate guidance and learning networks.
6- Pay attention to the influence of other factors in the school, the education system, the community and the culture that can support or that may hinder effective teaching.
7- Encourage evidence-informed teacher collaboration and self-reflection as strategies to enhance effectiveness and achieve consistency in improving effectiveness in all aspects of teaching.
8- Encourage monitoring and observation using appropriate research-based protocol to support professional learning and the development of effective practice among teachers and among subjects department.
9- Incorporate the students' perspectives and experiences to promote positive school and classroom climates that engage and motivate learners.

As a result of these research findings and the analysis of the data, the researcher recommends that the use of the value-added model be adopted as a useful tool to measure teacher effectivness.

REFERENCES

Benedict, A., Thomas, R., Kimerling, J. and Leko, C. (2013). Trends in Teacher Evaluation. *Teaching Exceptional Children*, 45(5), pp.60-68.

Braun, H. and Educational Testing, S. (2005). Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models. Policy Information Perspective. *Educational Testing Service*.

Culbertson, J. (2012). Putting the Value in Teacher Evaluation. *Phi Delta Kappan*, 94(3), pp.14-18.

Darling-Hammond, L. and Center for American, P. (2010). Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching. *Center for American Progress*.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. and Rothstein, J. (2012). Evaluating Teacher Evaluation. *Phi Delta Kappan*, 93(6), pp.8-15.

DeMitchell, T., DeMitchell, T. and Gagnon, D. (2012). TEACHER EFFECTIVENESS AND VALUE-ADDED MODELING: BUILDING A PATHWAY TO EDUCATIONAL MALPRACTICE?. *Brigham Young University Education & Law Journal*, [online] (2), pp.257-301.

Dossett, D. and Munoz, M. (2003). Classroom Accountability: A Value-Added Methodology.

Ellett, C. and Teddlie, C. (2003). Teacher Evaluation, Teacher Effectiveness and School Effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, [online] 17(1), pp.101-128.

Gallagher, M. (2012). How Principals Support Teacher Effectiveness. *Leadership*, [online] 41(3), pp.32-35.

Garrett, K. (2011). value added: do new teacher evaluation methods make the Grade?. *California Schools*, pp.P40-45.

Goe, L., Bell, C., Little, O. and National Comprehensive Center for Teacher, Q. (2008).

Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. *National Comprehensive Center for Teacher Quality*.

Goe, L., Croft, A. and National Comprehensive Center for Teacher, Q. (2009). Methods of Evaluating Teacher Effectiveness. Research-to-Practice Brief. *National Comprehensive Center for Teacher Quality.*

Hazi, H. and Rucinski, D. (2009). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *education policy analysis archives*, 17(5), pp.1-22.

Hsi-Chi, H. and Ming-Chao, L. (2010). A Study of Vocational High School Teacher's Professional Development Evaluation and Teaching Effectiveness. *International Journal of Learning*.

Islahi, F. and Nasreen, (2011). EVALUATION OF GENDER INFLUENCES ON THE EFFECTIVENESS OF SECONDARY SCHOOL TEACHERS. *International Journal of Education & Allied Sciences*.

Khda.gov.ae, (2014). *Government of Dubai* | *Knowledge and Human Development Authority* | *Home*.

Ko, J., Sammons, P., CfBT Education Trust (United, K., Hong Kong Institute of Education, (. and University of Oxford (England), D. (2013). Effective Teaching: A Review of Research and Evidence. *CfBT Education Trust*.

Little, O., Goe, L., Bell, C. and National Comprehensive Center for Teacher, Q. (2009). *A Practical Guide to Evaluating Teacher Effectiveness.*

Mangiante, E. (2010). Teachers matter: Measures of teacher effectiveness in low-income minority schools. *Educ Asse Eval Acc*, 23(1), pp.41-63.

McColskey, W., Egelson, P. and Southeastern Regional Vision for Education (SERVE), F. (1997). Designing Teacher Evaluation Systems That Support Professional Growth. Second Edition.

Milanowski, A. (2011). Strategic Measures of Teacher Performance. *Phi Delta Kappan*, 92(7), pp.19-25.

Murray, J. (2012). Development and Psychometric Evaluation of the Independent School Teacher Development Inventory. *The Journal of Experimental Education*, 80(3), pp.219-245.

Newton, X., Darling-Hammond, L., Haertel, E. and Thomas, E. (2010). Value-Added Modeling of Teacher Effectiveness: An Exploration of Stability across Models and Contexts. *Education Policy Analysis Archives*.

Paige, M. (2012). Using VAM in High-Stakes Employment Decisions. *Phi Delta Kappan*, 94(3), pp.29-32.

Piro, J., Wiemers, R. and Shutt, T. (2011). Using Student Achievement Data in Teacher and Principal Evaluations: A Policy Study. *International Journal of Educational Leadership Preparation*, 6(4), p.4.

Slater, H., Davies, N. and Burgess, S. (2011). Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England*. *Oxford Bulletin of Economics and Statistics*, 74(5), pp.629-645.

Stecher, B., Garet, M., Holtzman, D. and Hamilton, L. (2012). Implementing measures of teacher effectiveness. *Phi Delta Kappan*, 94(3), pp.39-43.

Yu, C. (2008). Book Review: Creswell, J., & Plano Clark, V. (2007). Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage. *Organizational Research Methods*, 12(4), pp.801-804.

Zaslavsky, O. and Leikin, R. (2004). Professional Development of Mathematics Teacher Educators: Growth Through Practice. *Journal of Mathematics Teacher Education*, 7(1), pp.5-32.

APPENDICES

Appendix 1: Student results in Y10 and Y11

Year 10 Biology

St. No.	Section	Gender	EOY/7	IGCSE/7
Section 1				
1	1	F	7	7
2	1	F	7	7
3	1	F	6	6
4	1	F	6	5
5	1	F	5	6
6	1	F	5	5
7	1	F	4	5
8	1	F	6	6
9	1	F	5	5
10	1	F	5	3
11	1	F	5	4
12	1	M	6	7
13	1	M	6	5
14	1	M	5	5
15	1	M	6	7
16	1	M	6	6
17	1	M	4	3
18	1	M	6	6
19	1	M	6	6
Section 2				
1	2	F	7	7
2	2	F	4	6
3	2	F	7	7
4	2	F	6	5
5	2	F	4	5
6	2	F	7	5
7	2	F	6	7
8	2	F	5	4
9	2	M	7	7
10	2	M	6	4
11	2	M	5	6
12	2	M	7	7

13	2	M	7	6
14	2	M	5	6
15	2	M	4	5
16	2	M	4	5
17	2	M	7	7

Year 10 Maths

St. No.	Section	Gender	EOY/7	IGCSE/7
Section 1				
1	1	F	7	6
2	1	F	7	7
3	1	F	7	7
4	1	F	6	5
5	1	F	6	5
6	1	F	7	6
7	1	F	6	4
8	1	F	5	4
9	1	F	5	6
10	1	F	5	6
11	1	M	5	6
12	1	M	6	5
13	1	M	5	6
14	1	M	7	6
15	1	M	7	7
16	1	M	7	6
17	1	M	7	6
18	1	M	6	6
19	1	M	5	5
20	1	M	5	4
21	1	M	5	6
22	1	M	4	4
23	1	M	4	3
Section 2				
1	2	F	7	5
2	2	F	7	6
3	2	F	7	5
4	2	F	7	6
5	2	F	6	6

6	2	F	6	6
7	2	F	6	6
8	2	F	6	4
9	2	F	5	5
10	2	F	5	6
11	2	M	7	6
12	2	M	5	4
13	2	M	7	7
14	2	M	7	5
15	2	M	7	6
16	2	M	7	7
17	2	M	7	6
18	2	M	6	6
19	2	M	6	5
20	2	M	6	7
21	2	M	6	6
22	2	M	5	4
23	2	M	5	5
24	2	M	5	5
25	2	M	5	6
26	2	M	4	6
27	2	M	4	4
Section 3				
1	3	F	5	5
2	3	F	6	6
3	3	F	7	5
4	3	F	7	7
5	3	F	6	7
6	3	F	6	6
7	3	F	6	6
8	3	F	6	6
9	3	F	6	7
10	3	F	4	6
11	3	F	4	4
12	3	M	6	6
13	3	M	7	6
14	3	M	7	7
15	3	M	6	6
16	3	M	6	6
17	3	M	6	5

18	3	M	5	5
19	3	M	5	5
20	3	M	5	6
21	3	M	4	4
Section 4				
1	4	F	7	7
2	4	F	6	6
3	4	F	7	7
4	4	F	6	7
5	4	F	6	5
6	4	F	6	7
7	4	F	6	6
8	4	F	5	5
9	4	F	5	3
10	4	F	5	4
11	4	F	4	6
12	4	M	7	7
13	4	M	4	4
14	4	M	6	5
15	4	M	7	6
16	4	M	7	7
17	4	M	7	6
18	4	M	6	5
19	4	M	6	6
20	4	M	6	6
21	4	M	5	6
22	4	M	5	4
23	4	M	5	6
24	4	M	4	6
25	4	M	4	3
Section 5				
1	5	F	7	7
2	5	F	5	6
3	5	F	7	6
4	5	F	5	4
5	5	F	5	4
6	5	F	5	3
7	5	F	5	6
8	5	F	5	6
9	5	M	7	7

10	5	M	5	4
11	5	M	6	7
12	5	M	7	7
13	5	M	7	7
14	5	M	6	6
15	5	M	6	6
16	5	M	6	5
17	5	M	6	5
18	5	M	6	6
19	5	M	6	7
20	5	M	5	5
21	5	M	5	3
22	5	M	5	3
23	5	M	5	6
Section 6				
1	6	F	7	7
2	6	F	6	6
3	6	F	6	6
4	6	F	6	5
5	6	F	6	5
6	6	F	6	5
7	6	F	5	5
8	6	F	5	5
9	6	F	5	4
10	6	F	5	4
11	6	M	7	7
12	6	M	7	7
13	6	M	7	7
14	6	M	7	7
15	6	M	7	6
16	6	M	6	7
17	6	M	6	5
18	6	M	5	6
19	6	M	4	4

Year 11 Biology

St. No.	Section	Gender	EOY/7	IGCSE/7
Section 1				
1	1	F	6	5

2	1	F	6	7
3	1	F	5	4
4	1	F	4	3
5	1	F	6	7
6	1	F	6	6
7	1	F	4	4
8	1	F	6	7
9	1	F	6	5
10	1	F	5	5
11	1	M	5	5
12	1	M	4	3
13	1	M	7	6
14	1	M	6	6
15	1	M	6	5
16	1	M	6	5
17	1	M	6	4
18	1	M	6	6
19	1	M	7	7
20	1	M	6	5

Year 11 Maths

St. No.	Section	Gender	EOY/7	IGCSE/7
Section 1				
1	1	F	6	5
2	1	F	6	6
3	1	F	6	6
4	1	F	6	6
5	1	F	5	5
6	1	F	5	2
7	1	F	7	6
8	1	F	6	4
9	1	F	6	6
10	1	F	5	5
11	1	M	6	6
12	1	M	7	6
13	1	M	7	5
14	1	M	6	5
15	1	M	6	5
16	1	M	6	5

17	1	M	6	3
18	1	M	6	4
19	1	M	6	6
20	1	M	5	0
21	1	M	5	5
22	1	M	5	5
23	1	M	7	6
24	1	M	5	0
Section 2				
1	2	F	7	5
2	2	F	4	4
3	2	F	7	6
4	2	F	5	4
5	2	F	7	6
6	2	F	7	6
7	2	F	5	4
8	2	M	6	6
9	2	M	6	6
10	2	M	6	6
11	2	M	7	6
12	2	M	7	6
13	2	M	7	5
14	2	M	7	6
15	2	M	6	4
16	2	M	6	5
17	2	M	6	0
18	2	M	5	3
19	2	M	6	4
20	2	M	7	5
21	2	M	6	0
22	2	M	5	4
Section 3				
1	3	F	7	6
2	3	F	5	4
3	3	F	7	5
4	3	F	7	6
5	3	F	7	6
6	3	F	6	4
7	3	F	6	4
8	3	F	5	5

9	3	F	5	4
10	3	F	7	6
11	3	F	7	5
12	3	F	7	5
13	3	F	7	6
14	3	M	7	6
15	3	M	7	6
16	3	M	7	6
17	3	M	7	5
18	3	M	6	4
19	3	M	6	0
20	3	M	6	4
21	3	M	6	4
22	3	M	5	4
23	3	M	5	4
24	3	M	5	3
25	3	M	5	5
26	3	M	5	2
27	3	M	7	7
28	3	M	6	6
29	3	M	6	5
30	3	M	6	5
31	3	M	5	3
32	3	M	5	3
33	3	M	5	5
34	3	M	4	4
35	3	M	5	3

Year 11 Physics

St. No.	Section	Gender	EOY/7	IGCSE/7
Section 1				
1	1	F	6	7
2	1	F	6	7
3	1	F	6	0
4	1	F	5	5
5	1	F	6	7
6	1	F	5	5
7	1	F	7	6

8	1	F	6	6
9	1	F	5	5
10	1	M	5	7
11	1	M	6	6
12	1	M	6	6
13	1	M	7	7
14	1	M	5	7
15	1	M	7	0
16	1	M	6	6
17	1	M	6	7
18	1	M	5	7
19	1	M	6	7
20	1	M	7	7
21	1	M	7	7
22	1	M	6	6
23	1	M	6	6
24	1	M	6	6
Section 2				
1	2	F	7	7
2	2	F	4	5
3	2	F	6	7
4	2	F	6	6
5	2	F	5	6
6	2	F	5	6
7	2	M	6	0
8	2	M	7	7
9	2	M	7	7
10	2	M	6	0
11	2	M	6	7
12	2	M	6	7
13	2	M	6	7
14	2	M	6	0
15	2	M	6	6
16	2	M	7	7
17	2	M	7	7
18	2	M	6	6
19	2	M	0	0
Section 3				
1	3	F	6	6
2	3	F	5	6

3	3	F	6	7
4	3	F	7	7
5	3	F	6	6
6	3	F	6	5
7	3	F	6	7
8	3	F	6	7
9	3	F	6	7
10	3	F	6	7
11	3	F	6	6
12	3	F	5	5
13	3	M	7	7
14	3	M	7	0
15	3	M	6	7
16	3	M	6	6
17	3	M	5	6
18	3	M	4	6
19	3	M	6	7
20	3	M	6	6
21	3	M	5	5
22	3	M	5	6
23	3	M	5	5
Section 4				
1	4	F	7	7
2	4	F	6	6
3	4	F	6	7
4	4	F	4	5
5	4	F	7	7
6	4	F	7	0
7	4	F	6	6
8	4	F	5	7
9	4	M	7	7
10	4	M	5	7
11	4	M	6	6
12	4	M	5	6
13	4	M	5	5
14	4	M	5	6
15	4	M	5	0
16	4	M	6	6
17	4	M	6	6
18	4	M	6	7

19	4	M	6	6
20	4	M	6	6
21	4	M	6	7
22	4	M	5	5
23	4	M	7	7
24	4	M	6	7
25	4	M	5	5

Appendix 2: School evaluation of the different teachers of different sections and subjects.

Subject Sections' Teachers	Gender of Section's Teacher		School Evaluation of Teachers
	Female	Male	
Y10 Biology Section 1	1	-	Good
Y10 Biology Section 2	1	-	Good
Y10 Math Section 1	1	-	Good
Y10 Math Section 2	-	1	Good
Y10 Math Section 3	1	-	Acceptable
Y10 Math Section 4	1	-	Good
Y10 Math Section 5	-	1	Good
Y10 Math Section 6	1	-	Good
Sub-Total Y10	**6**	**2**	**8**
Y11 Biology Section 1	1	-	Good
Y11 Math Section 1	-	1	Acceptable
Y11 Math Section 2	-	1	Good
Y11 Math Section 3	1	-	Acceptable
Y11 Physics Section 1	-	1	Acceptable
Y11 Physics Section 2	-	1	Acceptable
Y11 Physics Section 3	-	1	Acceptable
Y11 Physics Section (4)	1	-	Good
Sub-Total Y11	**3**	**5**	**8**
Total	**9**	**7**	**16**