

CORRELATING TEACHER SELF-ASSESSMENT SCORE WITH STUDENT ACHIEVEMENT IN ENGLISH AND MATHEMATICS

Muhammad Akram *

Sally J. Zepeda **

Abstract

This study aimed to measure the relationship between teacher self-assessment score and student achievement in English and Mathematics. The researchers developed a Self-assessment Instrument for Teacher Evaluation (SITE) based on five National Professional Standards for Teachers developed by the Ministry of Education, Pakistan. Using a convenient sampling method, 279 English or Mathematics teachers of grade 10 in 40 public boys' and girls' high schools in district Okara were surveyed who self-evaluated their performance on the SITE II. Additionally, based on the Lahore Board's annual matriculation examination results 2014, the achievement scores in English or Mathematics (n=7245) of students were also collected from teachers. The study found positive, weak or moderate, relationships between teacher evaluation scores and student achievement in English, as well as in Mathematics. The findings also revealed that Subject Matter Knowledge, Instructional Planning and Strategies, Assessment, Effective Communication, and Continuous Professional Development, individually, significantly predicted student achievement in English and Mathematics. The Subject Matter Knowledge, Instructional Planning and Strategies, and gender significantly explained 28% of the observed variance to predict student achievement in English. Only Subject matter Knowledge explained 12% of the variance in student achievement in mathematics.

Keywords: *Teacher evaluation, student achievement, self-assessment, national professional standards, professional development*

* Assistant Professor, University of Education, Lahore. makram@ue.edu.pk

** Professor, the University of Georgia, USA. szepeda@uga.edu

Introduction

This study investigated the relationship between teacher evaluation scores and student achievement in English and Mathematics in a sample of Pakistani public schools. Teacher evaluation is a formal and systematic process of examining teachers' performance (Stronge, 2006, 2010). The enduring purposes of teacher evaluation are to assess performance of educators not only for certification, tenure and promotion decisions, but also to support valid and legal decisions for termination, and to monitor changes in the performance to make improvements (Darling-Hammond, 1990; Peterson, 2000; Stronge, 2006, 2010; Zepeda, 2012; 2013).

Comprehensive teacher evaluation systems address *accountability* and *improvement* as two wide-ranging purposes which serve the needs of the individual as well as the community. *Accountability* is central to meet organizational objectives through summative evaluation, while *improvement* contributes to the professional development needs of individuals through formative evaluation (Sanders & Sullins, 2005; Stronge & Tucker, 1995). The combination of both— formative and summative evaluation—results in identifying and supporting effective teachers and also teachers who need targeted improvement (Zepeda, 2012).

Regardless of continent, teacher evaluation processes are vitally important to understand around the globe. Effective teachers demonstrate high levels of teaching expertise, meet accountability standards, and share professional knowledge with their colleagues (Hunt, Wiseman, & Touzel, 2009; Stronge & Tucker, 2000). Effective teachers care deeply about students and their success, and have distinctive qualities that characterize their effectiveness (Stronge & Tucker, 2000; Wright, Horn, & Sanders, 1997). Stronge and Tucker (2000) argued that effective teachers “absolutely, unequivocally, make a difference in student learning” (p. 1). Realizing that teachers matter, the researchers focused on examining the relationship between teacher evaluation scores and student achievement in English and Mathematics in Pakistan using a valid measure of teacher evaluation.

Teacher evaluation in Pakistan is perhaps the least focused area in its education system. In Pakistan, public school teachers are evaluated by the school administrators on the Performance Evaluation Report (PER). The PER is a generalized evaluation report focusing primarily on personality characteristics. Judging the personality characteristics of the individual is important from an ethical perspective but not from a teacher

quality point-of-view as “personality characteristics [do] not necessarily relate to the quality of teaching performance” (Shinkfield & Stufflebeam, 1995, p. 12). The United Nations Educational Scientific and Cultural Organization (UNESCO, 2006) reported that the performance appraisal system of teachers in Pakistan “is merely a formality...[as it] fails to provide any useful feedback or insight to a teacher’s performance” (p. 50).

To meet the challenges in teacher education in Pakistan, the Policy and Planning Wing of the Ministry of Education (MoE) implemented a Strengthening Teacher Education in Pakistan (STEP) project in collaboration with UNESCO in 2008. STEP focused on further developing the Professional Standards for Teachers in Pakistan. These efforts were the part of a larger international movement aimed to contribute to the educational quality as measures to impact positively student learning outcomes (MoE, 2009). The National Professional Standards led to teacher evaluation indicators such as Subject Matter Knowledge, Instructional Planning and Strategies, Assessment, Classroom Environment, and Continuous Professional Development. In 2008, these professional standards were formally adopted by all four provincial governments in Pakistan.

An important aspect of the Professional Standards is their alignment with Danielson’s Framework for Teaching (1996), Marzano’s causal teacher evaluation model domains (2011), and the eight standards summarized by Stronge (2010). Stronge developed performance appraisal rubrics for each standard with detailed descriptions, and many of these research-based standards are identical to the ones developed by the Ministry of Education, Pakistan. It was imperative, therefore, to identify which of these Professional Standards were highly effective in the Pakistani public school context. To our knowledge, perhaps no studies have examined National Professional Standards for Teacher evaluation in Pakistan since their implementation in 2008.

To fill this gap, we developed the Self-assessment Instrument for Teacher Evaluation (SITE II) for Pakistani public high school teachers. We investigated the relationship between teacher evaluation scores on SITE II and 10th graders’ achievement scores in English or Mathematics in the 2014 annual examination conducted by the Board of Intermediate and Secondary Education (BISE) Lahore.

Literature Review

This study investigated the relationship between teacher evaluation scores and student achievement of high school English and Mathematics in Pakistan. The study was informed by the literature on the teacher

evaluation in Pakistan, the research studies regarding teacher competency and teacher performance in Pakistan, Educational policies and National professional standards for teacher in Pakistan, and generally the literature on teacher self-assessment.

Teacher Evaluation in Pakistan

Since its birth in 1947, Pakistan has observed more than 15 education policy regimes directing educational improvement in the country (UNESCO, 2006). These policies focused on various aspects such as instituting a mechanism for teacher assessment (Kizilbash, 1998), increasing number of teachers, recruiting teachers, and improving the quality of teachers through better pre-service and in-service training, providing teachers professional development opportunities (MoE, 1998), and upgrading teacher qualifications (MoE, 2004). The provinces have been adopting these policies, to a limited extent, to improve the quality of education. In spite of the federal and provincial governments' policies, and a growing number of teacher training institutions that support these policies, the quality of teacher and teacher education in the public sector has been extremely low (UNESCO, 2006).

The issue of low quality of teacher performance was highlighted, for the first time, in the report of the National Commission on Education in 1959 (Government of Pakistan, 1959). Since then, teacher quality issues have been consistently addressed by the descendent policies until recently. In the latest analysis of teacher education in Pakistan, UNESCO (2006) reported:

Teacher education programs currently being run by the government institutes are not of the caliber to significantly raise the level of knowledge and skills of teachers to have any measurable impact in the students learning. The curriculum of these programs fails to develop in teachers the required pedagogical skills, subject knowledge, classroom delivery and questioning skills that would make these courses/programs worthwhile. (p. 44)

There is also a perceived consensus that the teacher evaluation in Pakistan is just a formality. Instead of using multiple data sources for measuring teacher quality as suggested by researchers (Peterson, 2000; Stronge, 2010), only one teacher evaluation report—Performance Evaluation Report (PER)—is being used in public high schools. The

purpose of the PER is to help authorities make decisions about the fitness for promoting employees (UNESCO, 2006). A concern of the PER is that the school administrators evaluate teachers' performance based on their own perceptions without data for variables such as teacher intelligence, emotional stability, will power, and knowledge of Islam. Research in the United States provides evidence that teachers' personal qualities and attitudes are not necessarily related to their performance (Shinkfield & Stufflebeam, 1995); this might be true in the Pakistani context as well where UNESCO (2006) reported that the current performance appraisal system of teachers fails to provide any useful feedback or insights to a teacher's performance.

Teacher Competence in Pakistan

The literature on teacher evaluation in Pakistan is insufficient; however, there exists a body of literature that reveals that the poor quality of teacher education has negatively impacted teacher quality in Pakistan (Almani, 2002). Almani developed Teacher's Self Performance Rating Scale (TSPRS) and compared the effects of in-service training on performance of secondary school teachers in the Hyderabad district in Pakistan. Almani found that teacher training significantly affected the classroom performance of female teachers as they performed better in various teacher quality indicators such as teaching methodology, teaching aids, communication style, classroom management, and evaluation. No statistically significant differences were found between male and female teachers in their content knowledge and classroom performance; however, male teachers rated themselves higher than female teachers on motivational techniques.

Bibi (2005) conducted a study to evaluate the personal and professional competencies of secondary school teachers in the Punjab province of Pakistan. Ten heads of teacher training institutions, 50 teacher trainers, 800 heads of secondary schools, and 4000 secondary school teachers were randomly selected as a sample. One questionnaire for each type of the sample was developed. The overall results revealed that a significant number of secondary teachers used grammatically incorrect language while teaching, and used ineffective teaching methods. Bibi found that the secondary school teachers did not have command over the subject they taught, had poor knowledge about audio visual aids, did not relate the lessons to daily life experiences, did not have test construction skills, and were unable to diagnose the learning difficulties of students.

Jumani (2007) examined the extent to which teachers trained through distant education possessed competencies in professional knowledge, communication, planning the teaching learning process, assessing student learning, reflecting, evaluating and planning for continuous improvement. Findings suggested that teachers focused on textual knowledge, did not adapt new concepts, and lacked the ability to structure curricular and co-curricular activities. Jumani found that teachers were neither competent of presenting subject matter, nor did they select appropriate instructional strategies. Jumani reported that teachers did not provide opportunities for students to apply knowledge, and they did not discuss students' performance issues with their parents.

Dilshad (2010) conducted a study to assess quality of teacher education in teacher training colleges and various departments of a public university in Pakistan. Dilshad surveyed 350 student teachers asking about teacher education preparation across various indicators such as quality of learning environment, content and outcomes. Dilshad found that low quality but lengthy course content, poorly equipped classrooms, use of English as the medium of instruction and lack of highly qualified teacher trainers were the main reasons for poor teacher training found in various institutions and university departments.

Aziz (2010) analyzed the effects of student demographic factors and teachers' competencies (Teaching Planning, Teaching Process, Classroom Management, Experience and Evaluating Techniques) on the achievement of secondary school students. Aziz sampled 60 heads of schools, 300 secondary school teachers and 1500 students through convenience sampling technique. Three questionnaires were developed; each for heads, teachers and students. Aziz found that 9th graders' achievement (pass percentage) was significantly correlated with teacher's scores on planning, classroom management, experience and evaluating techniques.

Based on these studies, the researchers have concluded that teachers were less competent in various teacher quality indicators, they had little knowledge of the content and audio visual aids, lacked in test construction skills, and had little knowledge of different teaching methodologies (Aziz, 2010; Bibi, 2005; Jumani, 2007).

Educational Policies and National Standards in Pakistan

Findings across these studies coupled with reports about poor teacher quality summarized by UNESCO, the Ministry of Education

(MoE) started a review process of education policies and five-year plans in 2005 to launch a new education policy. One major action plan was related to improving the quality of education by setting National Standards for educational inputs, processes, and outputs (MoE, 2009). These standards were designed to define competencies and skills deemed to be essential for teachers, to guide the detailed development of pre and in-service programs of teacher education, and to assure the public about the quality of their educators (MoE, 2009). Some of the most research-based teacher quality standards are discussed briefly.

Subject Matter Knowledge is a “teacher’s understanding of subject facts, concepts, principles, and the methods through which they are integrated cognitively to determine the teacher’s pedagogical thinking and decision making” (Stronge, 2010, p. 19). The research indicates that strong content knowledge of a teacher is positively associated with student learning, especially in Mathematics (Hill, Rowan, & Ball, 2005; Wenglinsky, 2002). Others found that subject matter knowledge shows small, statistically insignificant relationships with student achievement (Ashton & Crocker, 1987; Haney, Madaus, & Kreitzer, 1987).

Instructional Planning and Strategies are other important elements of measuring teacher quality and effectiveness. Effective teachers become supportive and persistent in keeping students on task, and they engage, motivate, and maintain students’ attention to their lessons (Stronge, 2007). The research indicates that teachers’ instruction and strategies have the most proximal relation with student learning (Marzano, 2011).

Assessment for learning is a process of evaluating student performance where the teachers gather, analyze, and use data to measure progress (Stronge, 2010). Assessment of student learning can be documented in various ways such as teacher observation, oral questioning, homework assignments, project products, student opinions, criterion-referenced tests, or norm-referenced tests (MoE, 2009; Stronge, 2010). Assessment which is aligned with learning targets, accompanied with frequent feedback, involves students, and documents what influences student learning (Black & William, 1998; Zacharias, 2007).

Students need an engaging *learning environment* to support their growth (Stronge, 2010). Effective teachers create an environment of respect and rapport in their classrooms by the ways, they interact with students and by the interaction, they encourage and cultivate among students (Danielson, 1996). Effective teachers focus on the organization of learning activities throughout teaching and learning, maximize instructional time, assume responsibility for student learning, and establish

rapport and trustworthiness with students by being fair, caring, and respectful (Good & Brophy, 1997; Wang, Haertel, & Walberg, 1994).

Effective Communication is an ability to deliver content meaningfully, create an engaging class culture, be sensitive to individual student needs, and connect with the student; first, as a person, and, then, as a learner (Cornett-DeVito & Worley, 2005). Stronge and Tucker (2003) stated effective teachers communicate effectively with students, model standard language, actively listen and respond in a constructive manner, establish and maintain multiple modes of communication between school and home. Effective teachers use knowledge of effective verbal, nonverbal, and written communication techniques and tools, and collaborate and support interactions with students and parents (MoE, 2009; Stronge, 2010).

Teacher Self-Assessment

Teacher self-evaluation is a frequently advocated data source for teacher evaluation (McGreal, 1983; Peterson, 2000). The self-assessment is a process in which teachers make judgments about the adequacy and effectiveness of their own knowledge, performance, and pedagogical skills for the purpose of self-improvement (Airasian & Gullickson, 2006). Research indicated that teachers do monitor and improve their own behavior in relation to goals and outcomes, act on self-gained data, and engage themselves in professional development activities (Peterson, 2000).

Teacher self-assessment makes teachers aware of their strengths and weaknesses, encourages collegial interactions and teacher development, assists in school improvement, and helps administrators in making decisions about teaching assignments (Peterson, 2000). Self-assessment gives teachers control over their own growth and treats teachers as professionals (Airasian & Gullickson, 2006). Research on self-assessment illustrates teachers, by themselves, are in many ways the best judges of their teaching performance and growth (Airasian & Gullickson, 2006; Stufflebeam & Shinkfield, 1985). Self-assessment evidence can provide support for what teachers do in the classroom and can present a fuller picture of teaching unobtainable from other sources (Berk, 2005). Also, teachers are more likely to act on self-assessment data than other resources (Centra, 1973). Collecting data through teachers' self-assessment is feasible, cost efficient, and time saving (Goe, Bell, & Little, 2008).

To date, we could not find any studies that used questionnaire methods that encompassed the National Standards for Teachers in Pakistan and measured teacher quality. To fill this gap, the researchers

developed a *Self-assessment Instrument for Teacher Evaluation* (SITE II) for Pakistani public high school teachers and used it to correlate 10th graders' achievement in English and Mathematics on the Lahore Board's annual examination 2012, in one district, Okara, province Punjab.

The following overall research questions were addressed in this study:

- What is the relationship between teacher evaluation score (on five performance evaluation scales as, Subject Matter Knowledge, Instructional Planning and Strategies, Assessment, Learning Environment, and Effective Communication) and student achievement in English and Mathematics?
- To what extent do the five scales measured through a self-assessment instrument and demographic variables relate to predict student performance in English or Mathematics in Pakistan?

Conceptual Framework

Teacher evaluation is a complex phenomenon that involves multifaceted procedures, aspects, and contexts. Therefore, it is hard to measure this phenomenon comprehensively through a single teacher evaluation model or theory. Through a careful review of various teacher evaluation models such as Danielson's Framework for Teaching (1996), Marzano's Causal Teacher Evaluation Model (2010), Stronge's (2010) work on teacher effectiveness standards, and the National Professional Standards for Teachers in Pakistan (2009), the researchers selected five standards which were compatible with the various standards-based teacher evaluation models in other countries, especially in the US. Based on the conceptual framework, it was assumed that five subscales of teacher evaluation would predict students' achievement, separately and combined, in English and / or Mathematics (see Figure 1).

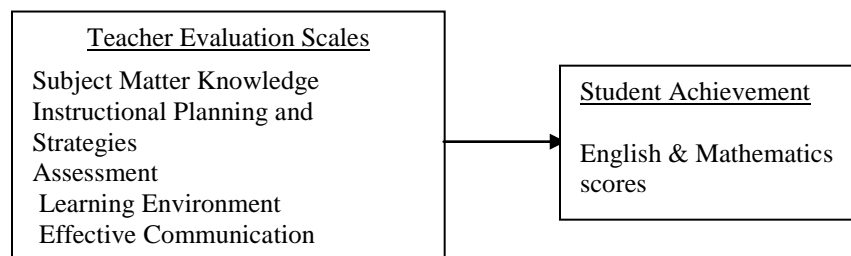


Figure 1. Conceptual Model of the Study

Methodology

Research Design

It was a correlational Study. Correlational research is a type of research in which the researcher correlates two or more variables. (Pearson r was used to correlate teacher self-assessment score with the mean score of his or her class students in English or Mathematics.

Participants

Data were collected through convenient sampling technique from 279 teachers across 40 public boys and girls high schools in district Okara. The teachers who taught English or Mathematics to 10th graders during the academic year of 2013-2014 were sample of the study. The response rate was 91%. Additionally, the 10th graders' achievement scores in English or Mathematics on the Board of Intermediate and Secondary Education (BISE) Lahore Annual Examinations 2014 were collected from each teacher ($n=7245$).

Instruments

Self-assessment Instrument for Teacher Evaluation (SITE II) developed by the authors and was used to collect data from 279 teachers and 7245 students of grade 10. SITE II was modified version of SITE (2012) which was influenced by work of Stronge (2010) who provided a detailed descriptions of the given standards, the key areas of standards, sample definitions of the standards, and what was supported by the seminal works of various researchers such as Aaronson et al. (2007), Good and Brophy (1997), Stronge and Tucker (2003), and Zacharias (2007). Stronge (2010) used teacher quality indicators which are “tangible behaviors that can be observed or documented to determine the degree to which a teacher is fulfilling” the particular standard (p. 23). SITE II was modified based on these teacher quality indicators. The response scales ranged from the lowest to the highest level of frequency of teacher quality such as Never (1), Rarely (2), Sometimes (3), Often (4), or Always (5).

Another type of data employed for this study was 10th graders' achievement scores in English and Mathematics earned by 7245 students in the Annual examination of the Board of Intermediate and Secondary Education (BISE) Lahore. The BISE Lahore is the responsible body for developing and administering tests of all school subjects for secondary school students. The tests of each subject are developed and marked by experienced teachers based on the rubrics provided by the BISE Lahore authorities. For this study, the teachers who self-evaluated on the SITE

were requested to provide the achievement scores of those students who were taught either English or Mathematics by them in grade 10 and who took BISE annual examination in 2014. The mean score of each class in English and Mathematics was calculated and used to correlate with each relevant teacher's score on the SITE. Since the Ministry of Education (2009) had communicated and handed over the draft of the national standards to the provincial authorities, we assumed that the teachers in the public school understood the standards and they accurately self-evaluated their performance.

Reliability and Validity

Initially, the content validity of the SITE was established through a panel of experts and a panel of practitioners. Both panels critically analyzed the content of the items and suggested minor modifications which were applied. Exploratory factor analysis using a varimax rotation was conducted to confirm whether 28 items represented the five scales of the teacher evaluation construct. The Eigen values larger than one were used to make decisions about the factor structure of the scales. Factor analysis confirmed five scales that explained almost 59% of observed variance in teacher evaluation scores. Item loadings of 28 items in five factors ranged from .79 to .43, demonstrating a solid adequacy for the construct. Confirmatory factor analysis was conducted using Lisrel 8.0. Chi-square index reported better fit, $\chi^2=643.14$, $p=0.0$. The confirmatory factor analysis revealed that the model represented SRMR=.05, GFI= .86. CFI=.98, RMSEA= .056, indicating that the measurement model fits the data well, and the model provides evidence of the construct validity.

Cronbach Alpha was calculated to measure the reliability of the teacher evaluation scale. The overall reliability of all items was high ($\alpha=.94$). The Cronbach alpha reliabilities of the scales were assessed as: subject matter knowledge (.89), Instructional Planning and Strategies (.86), Assessment (.83), Learning environment (.75), and effective communication (.73).

Data Analysis

Initially, the descriptive statistics of the five scales of teacher evaluation were calculated. Pearson r was calculated to measure the relationship of teacher evaluation score on self-assessment factors with

student achievement in English as well as in Mathematics. Multiple regression analysis was run to predict student achievement in English through teacher evaluation factors.

Results

Descriptive statistics reveal that teachers showed highest mean score on Subject Matter Knowledge (M=28.87, S.D. = 7.160), followed by Instructional Planning and Strategies (M=27.35, S.D. = 5.807), and Learning Environment (M=18.77, S.D. = 3.902). Teachers showed lowest mean score on Effective communication (M= 11.51, S.D. = 2.458) (see Table 1).

Table 1
Descriptive Statistics of the Teacher Evaluation Scales

	N	Min	Max	Mean	S.D.
Subject Matter Knowledge	279	8.00	40.00	28.87	7.160
Instructional Planning & Strategies	279	7.00	35.00	27.35	5.807
Assessment	279	5.00	25.00	18.67	3.850
Learning Environment	279	7.00	25.00	18.77	3.902
Effective Communication	279	3.00	15.00	11.51	2.458
Overall	279	6.00	28.00	20.96	.64

Pearson r was calculated to measure the relationship of teacher evaluation score on self-assessment factors with student achievement in English as well as in Mathematics. Each teacher's score on the SITE was used as a measure of teacher effectiveness which was, then, correlated with the mean score of the students of his or her class in English or Mathematics. The summary of the results is given in Table 2.

Table 2
Relationship between Teacher Evaluation score and Student Achievement in English and Mathematics

Factors	English	Mathematics
Subject Matter Knowledge	.50*	.34*
Instructional Planning and Strategies	.33*	.28*
Assessment	.25*	.24*
Learning Environment	.38*	.30*
Effective Communication	.16*	.17*

* $p < .05$

The results in Table 2 show that five factors of teacher self-assessment construct were significantly correlated with student achievement in English as well as in Mathematics. The Subject matter knowledge scale showed the highest correlation with student achievement in English ($r = .50$), and Mathematics ($r = .34$), followed by Learning environment ($r = .38$) in English and Mathematics ($r = .30$). Assessment was positively correlated with student achievement in English ($r = .25$) and Mathematics ($r = .24$). Effective communication showed least significant positive correlation with student achievement in English ($r = .16$) as well as in Mathematics ($r = .17$). In overall, weak to moderate positive significant correlations were found between teacher evaluation factors and student achievement in English and mathematics.

Before running multiple regression analysis, the researchers checked assumptions—linearity, constant variance, and outliers—and obtained scatterplots for each dimension. During eyeballing the scatterplots, the researchers observed that in most of the variables the data were scattered constantly, presenting a linear trend in each dimension. However, a couple of cases were lying away from the rest of the data. The researchers assumed these cases as potential outliers influencing the rest of the data. To confirm the influence, the researchers checked the Cook's D values of these cases which were quite similar to the rest of the values (less than .30). The Cook's D values confirmed that these cases were not influential outliers. Therefore, the researchers included all cases in preliminary model and ran further analysis.

Multiple regression analysis was run to predict student achievement in English through teacher evaluation factors. According to Table 3, Subject Matter Knowledge and Instructional Planning and Strategies, significantly predicted student achievement in English, $F(3, 275) = 34.826, p < .001, R^2 = .28$. The r^2 value in the teacher evaluation model showed that 28% of the observed variance in student achievement in English could be explained through Subject Matter Knowledge and Instructional Planning and Strategies, and teacher gender. The results indicated that female teachers with higher teacher evaluation scores on Subject Matter Knowledge and Instructional Planning and Strategies were expected to have higher achievement in English than male teachers. The rest of the predictors, including teacher experience, did not significantly contribute to the student achievement in English. Model summary is given in Table 3.

Table 3
Teacher Evaluation Model for Student Achievement in English

Parameter	Parameters (b)	Standardized estimate (Beta)	<i>t</i>	<i>p</i>
Subject Matter Knowledge	.503	.565	9.494	<.001
Instructional Planning and Strategies	-.323	-.125	-2.094	<.001
Teacher Gender	1.324	.103	2.001	<.05

Further, multiple regression analysis was conducted to predict student achievement in Mathematics based on teachers self-assessment scores. (See table 4).

Table 4
Teacher Evaluation Model for Student Achievement in Mathematics

Parameter	Parameters (b)	Standardized estimate (Beta)	<i>t</i>	<i>p</i>
Subject Matter Knowledge	.268	.342	6.049	<.001

According to Table 4, Subject Matter Knowledge significantly predicted student achievement in Mathematics, $F(1, 277) = 35.588$, $p < .001$, $R^2 = .12$. The r^2 value in the teacher evaluation model showed that 12% of the observed variance in student achievement in Mathematics could be explained through Subject Matter Knowledge. The rest of the predictors, including teacher experience, did not significantly contribute to the student achievement in Mathematics.

The purpose of this study was to measure the relationship between teacher self-assessment evaluation score and their students' achievement in English and Mathematics at the secondary level. The study found that subject matter knowledge, instructional planning and strategies, assessment, classroom environment, and communication significantly correlated with student's achievement in English as well as mathematics. The study found that teacher subject matter knowledge, instructional planning and strategies, and teacher gender significantly predicted 28% of variance in student achievement in English. The study also found that subject matter knowledge of teachers significantly predicted 12% of variance in student achievement in Mathematics. Teacher experience did not predict student achievement in English as well as in Mathematics.

The Self-assessment Instrument for Teacher Evaluation provided evidences of high validity and reliability.

Discussion

This study focused on measuring the relationship between teacher evaluation score based on self-assessment and student achievement in English and Mathematics. The study found five teacher performance evaluation indicators measured through the SITE II to be significantly correlated with student achievement in English as well as in Mathematics. These findings were consistent with the previous research (Gallagher, 2004; Kimball et al., 2004; Milanowski, 2004; Odden, 2004). The factor-wise correlation between teacher evaluation scores showed smaller or moderate positive relationships with student achievement in English ($r=.16$ to $.50$) as well as in Mathematics ($r=.17$ to $.34$). Marzano (2010) found similar results from Oklahoma (phase I) that 5 of 9 essential indicators (such as classroom assessment, instruction, & professional development) of teacher quality significantly showed small or moderate relationship with student achievement in [English] reading ($r=.33$ to $.53$) and Mathematics ($r=.31$ to $.39$); at Phase II, however, 6 of 9 correlations were significant for reading ($r=.11$ to $.40$), and only 1 for Mathematics ($r=.04$ to $.40$).

Teacher experience did not show significant relationship with student achievement in English as well as in Mathematics. This finding is similar to the findings of the studies conducted by Gallagher (2004) and White (2004) based on Danielson's Framework for teaching (1996). The researcher used teacher experience as a continuous scale and did not make categories because converting a continuous variable into a categorical variable can cause considerable loss of important information in regression analysis (Royston, Altman, & Sauerbrei, 2005).

The small or moderate positive relationships between teacher evaluation and student achievement have been found in previous studies, so they are not unexpected. Milanowski (2004), for example, argued on relatively small correlations (.3 to .4) between teacher evaluation and student achievement by stating:

It is important to recognize that high correlations between teacher evaluation scores and student achievement scores are unlikely to be found for reasons including error in measuring teacher performance, error in measuring student performance, lack of alignment between the curriculum taught by teachers and the student tests, and the role of student motivation and related characteristics in producing student learning. (p. 50)

Therefore, comparing the results of the current study with previous research, it can be concluded that the findings of this study confirmed that the SITE II provided a consistent, valid, and reliable evidence of

teacher quality measures. The findings of this study are encouraging and they contribute to better understanding of the complex teacher effectiveness construct through individual as well as collective lens. This preliminary study provided evidence of the reliability of the SITE II used to measure teacher quality. The SITE II helped produce a significant level of variance among teachers' scores on some of the teacher quality indicators. The results also confirmed the theoretical assertion as a means of finding that effective teachers in Pakistani public schools not only demonstrate higher levels of performance on certain teacher quality indicators, but also they show higher levels of student achievement in English as well as in mathematics. The work of Stronge (2010) has potential application of the complex teacher evaluation construct measured through a self-assessment tool.

The findings provide initial evidence of the effectiveness of the teacher quality indicators as designed by the Ministry of Education, Pakistan.

Conclusion

The findings of this study are encouraging. In this preliminary study, we found some evidence of the potential use of the SITE II used for measuring teacher quality. The study found significant relationship between teacher evaluation scores and student achievement. Though the researchers used the SITE II for English as well as Mathematics teachers, English teachers, perhaps, found the SITE more interesting and useful than Mathematics teachers. This study involved some limitations such smaller sample size of teachers, non-randomized sampling technique, and use of English language in the SITE II instead of Urdu which is the national language of Pakistanis. If the SITE II is translated into Urdu and tested over a large number of teachers, especially the mathematics teachers, selected through a randomized technique, the results might be different from the findings of this study. Also, perhaps it is reasonable to think of testing the SITE II on the teachers in other subject areas and varied grade levels; it might help identify any causes related to the language proficiency or other problems. Last, but not least, making summative evaluation judgments based on these results requires attention of the policymakers as student achievement can be affected by various factors such as private tuition, socio-economic status and so on. Based on weak to moderate relationships and the limitations, the researchers suggest that any generalizations should be made cautiously.

Recommendations

The provincial as well as the district governments, however, should ensure that all teachers are provided with the complete document of the National Standards because teachers must have deeper understandings of these Professional Standards in relation to their effectiveness and impact on student learning. Since teachers have never self-evaluated before, initially, the policymakers and district authorities may adopt and introduce the SITE in public schools, and take initial steps toward implementing these Standards.

REFERENCES

- Airasian, P. W., & Gullickson, A. (2006). Teacher self-evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practices* (2nd ed., pp. 186-211). Thousand Oaks, CA: Corwin Press.
- Almani, A. S. (2002). *A comparative study of effects of in-service training and the performance of secondary school teachers* (Doctoral dissertation). Retrieved from <http://pr.hec.gov.pk/Thesis/2739H.pdf>
- Ashton, P., & Crocker, L. (1987). Systematic study of planned variations: The essential focus of teacher education reform. *Journal of Teacher Education, 38*(3), 2-8. doi:10.1177/002248718703800302
- Aziz, M. A. (2010). *Effects of demographic factors and teachers' competencies on the achievement of secondary school students in Punjab* (Doctoral dissertation). Retrieved from <http://pr.hec.gov.pk/Thesis/727S.pdf>
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education, 17*(1), 48-62. Retrieved from <http://www.isetl.org/ijtlhe/pdf/IJTLHE8.pdf>
- Bibi, S. (2005). *Evaluation study of competencies of secondary school teachers in Punjab* (Doctoral dissertation). Retrieved from <http://pr.hec.gov.pk/Thesis/377.pdf>
- Black, P. J., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*(1), 7-73. doi:10.1080/0969595980050102
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 329-375). New York, NY: Macmillan.
- Centra, J. A. (1973). Self-ratings of college teachers: A comparison with student ratings. *Journal of Educational Measurement, 10*(4), 287-295. doi:10.1111/j.1745-3984.1973.tb00806.x

- Cornett-DeVito, M., & Worley, D. W. (2005). A front row seat: A phenomenological investigation of students with learning disabilities. *Communication Education, 54*(4), 312-333. doi:10.1080/03634520500442178
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (1990). Teaching and knowledge: Policy issues posed by alternate certification for teachers. *Peabody Journal of Education, 67*(3), 123-154. doi:10.1080/01619569009538694
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research, 53*(3), 285-328. doi:10.3102/00346543053003285
- Dilshad, R. M. (2010). Assessing Quality of Teacher Education: A Student Perspective. *Pakistan Journal of Social Sciences, 30*(1), 85-97. Retrieved from http://www.bzu.edu.pk/PJSS/Vol30No12010/Final_PJSS-30-1-08.pdf
- Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79*(4), 79-107. doi:10.1207/s15327930pje7904_5
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Retrieved from National Comprehensive Center for Teacher Quality website: <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>
- Good, T. L., & Brophy, J. E. (1997). *Looking in classrooms* (7th ed.). New York, NY: Longman.
- Government of Pakistan (1959). Report of the Commission on national Education.
- Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms Talismanic: Testing teachers for the improvement of American education. In E. Z. Rothkopf (Ed.), *Review of Research in Education, 14*, 169-238. doi:10.3102/0091732X014001169
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406. doi:10.3102/00028312042002371
- Hunt, G. H., Wiseman, D. G., & Touzel, T. J. (2009). *Effective teaching: Preparation and Implementation* (4th ed.). Springfield, IL: Charles C Thomas Publishers.
- Jumani, N. B. (2007). *Study on the competencies of the teachers trained through distance education in Pakistan*. (Unpublished post doctoral research). Retrieved from http://eprints.hec.gov.pk/3517/1/DR_NABI_BUX_JUMANI_post_doc_rport.pdf
- Kauchak, D., Peterson, K., & Driscoll, A. (1985). An interview study of teachers attitudes toward teacher evaluation practices. *Journal of Research and Development in Education, 19*(1), 32-37.
- Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment

- results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78. doi:10.1207/s15327930pje7904_4
- Kizilbash, H. H. (1998). Teaching teachers to teach. In P. Hoodbhoy (Ed.), *Education and the state: Fifty years of Pakistan* (102-135). New York, NY: Oxford University Press.
- Kunter, M., Baumert, J., & Koller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494-509. doi:10.1016/j.learninstruc.2007.09.002
- Marzano, R. J. (2011). *Research base and validation studies on the Marzano evaluation model*. Retrieved from
- McGreal, T. L. (1983). *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Messick, S. (1989). Validity. In R. L. Linn (3rd Ed.), *Educational measurement* (pp. 13-103). New York, NY: Macmillan.
- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33-53. doi:10.1207/s15327930pje7904_3
- Ministry of Education. (1998). *National education policy 1998-2010*. Retrieved from <http://planipolis.iiep.unesco.org/upload/Pakistan/Pakistan20Educational20Policy201998-2010.pdf>.
- Ministry of Education. (2004). *Education sector reform: Action Plan (2001-02—2005-06)*. Retrieved from <http://planipolis.iiep.unesco.org/upload/Pakistan/Pakistan20Education20Sector20Reform202002-2006.pdf>
- Ministry of Education. (2009). *National education policy 2009*. Retrieved from http://planipolis.iiep.unesco.org/upload/Pakistan/Pakistan_National_education_policy_2009.pdf
- Ministry of Education. (2009). *National professional standards for teachers in Pakistan*. Retrieved from <http://unesco.org.pk/education/teachereducation/files/National20Professional20Standards20for20Teachers.pdf>
- Odden, A. (2004). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education*, 79(4), 126-137. doi:10.1207/s15327930pje7904_7
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Royston, P., Altman, D. G., Sauerbrei, W. (2005). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, 25, 127-141. doi:10.1002/sim.2331
- Sanders, J. R., & Sullins, C. D. (2005). *Evaluating school programs: An educator's guide*. Thousand Oaks, CA: Corwin Press.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher evaluation: Guide to effective practice*. Norwell, MA: Kluwer Academic Publishers.

- Stronge, J. H. (2006). Teacher evaluation and school improvement. In J. H. Stronge (2nd ed.), *Evaluating Teaching: A guide to current thinking and best practice* (pp. 1-22). Thousand Oaks, CA: Corwin Press.
- Stronge, J. H. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Stronge, J. H. (2010). *Evaluating what good teachers do: Eight research-based standards for assessing teacher excellence*. Larchmont, NY: Eye on Education.
- Stronge, J. H., & Tucker, P. D. (1995). Performance evaluation of professional support Personnel: A survey of the states. *Journal of Personnel Evaluation in Education*, 9(2), 123-137. doi:10.1007/BF00972655.
- Stronge, J. H., & Tucker, P. D. (2000). *Teacher evaluation and student achievement*. Washington, DC: National Education Association.
- Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. Larchmont, NY: Eye on Education.
- United Nations Educational Scientific and Cultural Organization. (2006). *Situation analysis of teacher education: Towards a strategic framework for teacher education and professional development*. Retrieved from <http://unesco.org.pk/education/documents/step/SituationAnalysis-StrategicFrameworkforTeacherEducation.pdf>
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1994). What helps student learn. *Educational Leadership*, 51(4), 74-79. Retrieved from <http://www.casdk12.net/GHS04/SRB/5-Curriculum/What20Helps20 Students20Learn.pdf>
- Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12), 1-30. <http://epaa.asu.edu/ojs/article/view/291/417>
- White, B. (2004). *The relationship between teacher evaluation scores and student achievement: Evidence from Coventry, RI*. Retrieved from University of Wisconsin-Madison, Consortium for Policy Research in Education website: <http://cpre.wceruw.org/papers/CoventryAERA04.pdf>
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teachers and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57-67. doi:10.1023/A:1007999204543
- Zacharias, N. T. (2007). Teacher and student attitudes toward teacher feedback. *RELC Journal*, 38(1), 38-52. doi:10.1177/0033688206076157.
- Zepeda, S. J. (2012). *Instructional supervision: Applying tools and concepts* (3rd ed.). New York, NY: Routledge
- Zepeda, S.J. (2013). *The principal as instructional leader: A practical handbook* (3rd ed.). New York, NY: Routledge.

Received on: June 30, 2015

Revised on: November 25, 2015

Accepted on: December 17, 2015