

Development and Validation of a Teacher Self-assessment Instrument

¹ *Muhammad Akram*, ² *Sally J. Zepeda*

¹ *University of Education, Lahore, Pakistan*

² *University of Georgia, USA*

(Email: makram@ue.edu.pk)

The purpose of this study was to develop and validate a *Self-assessment Instrument for Teacher Evaluation (SITE II)* based on five National Professional Standards for Teachers developed by the Ministry of Education, Pakistan: subject matter knowledge, instructional planning and strategies, assessment, learning environment, and effective communication. The data were collected from 279 English and mathematics teachers of grade 10 in 40 public boys' and girls' high schools in district Okara who self-evaluated their performance on five teacher performance components. The overall reliability of the questionnaire was found high ($\alpha = .94$). The SITE II factor structure was discovered through exploratory factor analysis. Confirmatory factor analyses provided evidence of construct validity of the questionnaire. Significant positive relationship was found between teachers' scores on self-evaluation questionnaire and their students' achievement ($n = 7245$) in English as well as in mathematics. The findings suggest that the questionnaire is valid and efficient tool for measuring components of teacher self-assessment. Teachers can use this scale for evaluation of their own teaching and take remedial actions. Mentors and supervisors can use it for diagnostic purpose and designing professional development courses for teachers.

Keywords: *teacher evaluation, student achievement, self-assessment, performance evaluation report, national professional standards*

Introduction

Teacher evaluation is a formal and systematic process of examining teacher performance (Stronge, 2010). Under the era of accountability, when teaching standards have been set (Government-of-Pakistan, 2009b) and teachers are required to perform effectively to meet the standards, evaluating teachers to identify effective and ineffective teachers is a vitally important process (Ngoma, 2011). Effective teachers are expected to demonstrate competence in subject matter, perform high levels of teaching skills, meet the accountability standards, share professional knowledge with their colleagues, care deeply about students and their success, and hold distinctive qualities that characterize their effectiveness.

Previously, various efforts have been made around the globe to identify effective teachers (Korthagen, 2004). In the USA, several researchers and institutions have developed rigorous teacher

evaluation frameworks as a basis for developing rubrics for teacher evaluation (Ingvarson, 2002). Danielson's Framework for Teaching (Danielson, 1996), The TAP: System for Teacher and Student Achievement (1999), The Bill and Melinda Gates Foundation's Measures of Effective Teaching (2009), and Marzano's Causal Teacher Evaluation Model (2010) are well-known examples of such frameworks. These frameworks have been extensively used in various teacher evaluation systems to evaluate teacher performance (Gallagher, 2004; Kimball, White, Milanowski, & Borman, 2004; Milanowski, 2004).

The evaluation can base on self-assessment or assessment of external evaluators. In informal settings principals evaluate teachers' performance on such rubrics (Boulter, 1987). Many authors have reported two problems when principals rate teachers. Firstly they tend to be lenient (Kauchak, Peterson, & Driscoll, 1985; Peterson, 2000), and secondly they base their ratings on limited number of observations of short durations (Zepeda, 2014).

In Pakistan, Performance Evaluation Report (PER) is used to evaluate teachers' performance in public schools and head of the school evaluates teachers (Akram, 2012). The most part of this report comprised personality characteristics such as religious affiliation and honesty which do not necessarily demonstrate teacher competence (Laking, 2007). Besides PER, various authors such as Almani (2002), Aziz (2010), Bibi (2005), Dilshad (2010) and Jumani (2007) measured teacher competencies through teacher competency questionnaires which either lack.

Evaluating teacher performance through just one source (PER) is problematic as the researchers are of the view that multiple data sources must be used to evaluate teacher performance (Peterson, 2000, Stronge, 2006; 2010; Darling-Hammond, 2011). Peterson (2006) stated no single data source such as student achievement, or client survey is valid and feasible for each and every teacher and no single data source works well for evaluating overall performance of a teacher. Multiple data sources such as self-evaluation (Airasian & Gullickson, 2006), student surveys (Stronge & Ostrander, 2006), Portfolio (Wolf, 2006), multiple observations (Zepeda, 2006), and artifacts provide comprehensive and clear picture of teacher performance (Peterson, 2004).

Given the weak evidence of teacher quality indicators in Pakistan, it was imperative to address two important purposes: (1) to search for teacher quality indicators compatible with the international teacher quality standards, and (2) to search for other data sources such as teacher evaluation tool(s) other than the principal's ratings. The first purpose was addressed by employing the National Professional Standards for Teachers in Pakistan developed by the Ministry of Education (2009). These National Professional Standards comprise important teacher evaluation indicators such as Subject Matter Knowledge, Instructional Planning and Strategies, Assessment, Classroom Environment, communication, and others which are highly compatible with the international standards of teacher quality. These professional standards have been formally adopted by all four provincial governments in Pakistan. To address the second purpose mentioned above, the researchers searched literature for developing two questionnaires: Self-

assessment Instrument for Teacher Evaluation (SITE), and Students' Perceptions of Teacher Effectiveness Questionnaire (SPTEQ). The scope of this study is limited to the development and validation of the SITE only.

SITE was initially developed earlier that included 6 factors of Teacher self-assessment (Akram, 2012). However, some of the issues related to reliability were attached with the SITE. Though SITE, in overall, was reliable ($r = .84$), however, the subscales of the SITE showed relatively low level of Cronbach Alpha (between .60 and .70). Nunnally (1978) stated that the reliability of the scale having Cronbach Alpha more than .70 is "acceptable" and less than .70 is "questionable". Moreover, professional development indicator seems to be less relevant for teacher self-evaluation as teachers do not have broader concept of professionally development where teachers can develop themselves professionally through book readings, mentoring, using portfolios or through peer coaching etc. Teachers believe that workshops and training sessions are the true sources of professional development (Khan & Begum, 2012); however every teacher does not have necessarily access to such opportunities. That is why professional development indicator was excluded from the SITE II. Also, some of the items in Assessment and Effective communication factors did not perform well. Sample size was also relatively smaller ($n=155$). Keeping in view these issues, SITE was modified and reproduced in the form of SITE II and used with relatively larger sample size.

The researchers hoped this exploratory study would provide initial data-based evidence of the effectiveness of the National Professional Standards in Pakistani context. Additionally, the SITE might be used in American schools as an alternative to evaluators' ratings which have been shown to be lenient, flawed, and biased (Kauchak et al., 1985; Milanowski, 2004; Peterson, 2000).

Conceptual Framework

This study has been theorized on the self-directed learning approach described by the humanistic theorists. The humanistic school believes that emotional factors, and personal growth and development, are the highest values, and it

argues that these are ignored in a society which is unduly materialistic, objective and mechanistic. Humanistic psychologists believe that learners should be allowed to pursue their own interests and talents in order to develop themselves as fully as possible. That can be best achieved through self-directed learning. Petty (2009) developed self-directed model based on humanistic principles to encourage the individuals to develop and to improve. Petty Described that self-directed learning is a cyclic process that starts from self-evaluation where one reflects and evaluates his or her knowledge and skills. Based on the reflections made on the self-evaluation, the individuals set goals and targets for future improvement. Goal setting leads to devising action plan to bridge the gap between current and required performance. And finally, action is taken by the individual to implement the action plan.

Self-assessment is conducted in nonthreatening situation where the teachers do not have fear of bad evaluations that can affect their promotion, salary increment, or other benefits related to job. Self-assessment can be related to one's judgment about one's effectiveness and adequacy of the knowledge, performance, or beliefs. In that case, teacher self-evaluation can be viewed as a teachers' judgment about his or her effectiveness of the content knowledge, classroom teaching, using effective teaching strategies, assessing students' performance, having effective communication with their students and so on. The teachers are the best judge of their own performance as they are capable of assuming responsibility for much of their own professional development given times, encouragement, and resources (Peterson & Comeaux, 1990). The researchers used Petty's (2009) model as a theoretical framework of this study.

National Professional Standards for Teachers in Pakistan

To meet the challenges faced in the field of 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 the United Nations Educational Scientific and Cultural Organization (UNESCO) in 2008 (Lister, Bano, Carr-Hill, & MacAuslan, 2010). The STEP project

basically focused on developing the Professional Standards for Teachers in Pakistan in consultation with stakeholders in the country. This step was taken as a part of a larger international movement of quality assurance that contributes to the educational quality and impacts student learning outcomes in various fields of human endeavor (Ministry of Education, 2009).

One of the fundamental elements of teacher attributes that contribute to student learning and achievement is a teacher's knowledge of the subject matter (Danielson, 1996; Stronge, 2010). The subject matter knowledge refers to the amount and organization of knowledge (Shulman, 1986). Subject Matter Knowledge is related to teacher's understanding of subject information, concepts, principles, and pedagogical thinking and decision making (Stronge (2010). Effective teacher effectively addresses the appropriate curriculum standards, and integrates key elements and higher-level thinking skills in instruction (Danielson, 1996; Stronge, 2010). An effective teacher demonstrates accurate knowledge of the subject matter, links previous knowledge with the current learning experiences, demonstrates the skills relevant to the subject areas, and understands developmental needs of the age groups (Stronge, 2010). The research indicates that strong content knowledge of a teacher is positively associated with student learning, especially in mathematics (Aaronson, Barrow, & Sander, 2007; Hill, Rowan, & Ball, 2005). Others found, however, that the subject matter knowledge shows small, statistically insignificant relationships, both positive and negative (Quirk, Witten, & Weinberg, 1973).

Instructional planning and Strategies, another important element of measuring teacher effectiveness requires an effective teacher to use varying instructional strategies and techniques to maximize student learning (Stronge, 2010; Tomlinson, 1999). Shulman (1986) stated that effective teachers are required to demonstrate strategies most likely to be fruitful in reorganizing the understanding of the learners. Effective teachers also 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 (Cohen, Raudenbush, & Ball, 2003; Marzano, 2007; Walberg, 1984). (Marzano et al. (2011) developed various instructional strategies in over 300 experimental studies to investigate the relationship of instructional strategies with student achievement and found that student achievement increased by 16 percentile points when teachers used instructional strategies. Various other studies also found similar results (Tomlinson, 1999; Walberg, 1984).

Assessment for learning is a process of evaluating student performance where the teacher gathers, analyzes, and uses data to measure learners' progress (Stronge, 2010). Student assessment provides an overview of what the teacher has taught to the students. Assessment provides diagnostic information regarding students' mental readiness for learning new content, provides formative and summative information needed to monitor student progress, helps keep student motivated, helps students accountable for their own learning, and helps students retain what they have learned (Sanders, 2000).

Stronge (2010) giving the examples of effective teachers, stated that they use assessment data to develop expectations for students, use a variety of formal and informal assessment strategies, collect and maintain record of student assessment, and develop tools that help students assess their own learning needs. Research indicates that assessment positively influences student learning (Stronge, 2010). Assessment which is aligned with learning targets, accompanied with frequent feedback, involves students deeply in classroom, and is documented properly through record keeping influences student learning (Black & Wiliam, 1998). The researchers found that formative assessment shows positive effect on student achievement, especially with low achievers (Black & Wiliam, 1998).

Students need an engaging and stimulation learning environment to support student growth (Stronge, 2010). Effective teachers create respect their students, interact with them, and cultivate environment conducive to learning for 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 (Frey & Schmitt, 2007; Marzano, Pickering, & McTighe, 1993).

Research indicates that in a positive learning environment, effective teachers develop functional floor plans and material placement for optimal benefit, and establish classroom rules and procedures (Evertson, 1985; Stronge, 2007). Kunter, Baumert, and Koller (2007) found that the students' perceptions of rule clarity and teacher monitoring are positively related to their development of academic interest in secondary school mathematics classes. Effective teachers have less disruptive student behaviors than do less effective teachers (Stronge, Ward, Tucker, & Hindman, 2007). Wang, Wang, Wang, and Huang (2006) found that classroom instruction and climate significantly affect student aptitude. Summarizing to these findings, a positive classroom environment increases student-teacher interaction, maximizes instructional time, and helps students improve their achievement.

The ability to communicate is yet another essential requisite for teacher effectiveness (Fullan, 1993). Stronge and Tucker (2003) stated that 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, and follow the school policies regarding communication of student information. Effective teachers use knowledge of effective verbal, nonverbal, and written communication techniques and tools, and collaborate and support interactions with students and parents (Government-of-Pakistan, 2009a). Effective teachers explain concepts in simple and logical sequence, and explain lessons according to the age and ability of the students (Stronge, 2010). Catt, Miller, and Schallenkamp (2007) encouraged an open, warm, and communicated environment that invited students' comments. The results of the Catt et al. (2007) study revealed that open and warm communication with the students, parents, and community helped teachers as well as students perform better. These findings show that effective teachers can maximize student learning though discussing students'

problems with their colleagues, and adapt those behaviors followed by the teachers better in communicating with students.

Teacher Self-Assessment

Why the researchers argue in favor of using self-assessment instrument for teacher evaluation is based on the literature that supports the idea of using self-assessment as an opportunity for one's self-improvement and professional development (Centra, 1973, 1977; Peterson, 2000). Self-assessment is a very powerful tool for measuring teacher quality as side by side using the ratings done by principals or other administrators (Danielson, 1996; Peterson, 2000). Principals or administrators judge teachers' performance through observation and complete ratings or checklists during observation process (Darling-Hammond, Wise, & Pease, 1983; Medley & Coker, 1987). Rating teachers on the basis of limited observations and then generalizing those ratings over their overall teaching performance provides limited evidence of reliability (Zepeda, 2014). It is quite possible that during those observations teachers were well prepared and demonstrated high performance, or they were stuck with some serious social problems and demonstrated very low or average performance. Supervisors, therefore, can only capture limited sample of teachers' teaching performance through observation (Zepeda, 2014).

Studies show that supervisor evaluations are often influenced by a number of non-performance factors such as the age and gender of the supervisor and subordinate and the likability of the subordinate (Bolino & Turnley, 2003; Heneman, Greenberger, & Anonyuo, 1989; Varma & Stroh, 2001). Moreover, principals are generally effective at identifying the best and the worst teachers but not able to distinguish teachers in the middle of the achievement distribution (Medley & Coker, 1987). Further, supervisors are vulnerable to teachers' reactions in terms of subject matter expertise, school context, peer evaluation, use of portfolio, evaluator's competency, strictness, and leniency in ratings (Milanowski & Heneman, 2001). Teacher self-evaluation, on the other hand, 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. Research indicated that teachers do monitor and improve their own behavior in relation to goals, expectations, and outcomes, act on self-gained data, and engage themselves in professional development activities (Festinger, 1954; 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. As demonstrated by some of the studies, teachers, by themselves, are the best judges of their teaching performance and growth (Clandinin & Connelly, 1988). Danielson (1996) recommended self-evaluation as the most powerful tool of measuring teacher quality. Though there is possibility that experienced teachers would rate themselves higher on teaching effectiveness indicators as Almani (2002) found in his study, a self-assessment evidence can provide support for what teachers do in the classroom and can present a picture of their teaching unobtainable from any other sources (Berk, 2005). Also, teachers are more likely to act on self-gained data than on information from other resources (Centra, 1973). Moreover, teachers' perceptions would be based on multiple data sources such as samples of students' work, logs of professional development activities, and contacts with families which are important elements of the teacher quality indicators. Lastly, collecting data through teachers' self-assessments is feasible, cost efficient, and time saving. The researchers, therefore, developed and then used the Self-Assessment Instrument for Teacher Evaluation (SITE II) as a single method of data collection for this study. The researchers hope that the self-assessment instrument might serve as an alternative to the ratings of principals and school administrators.

Research Design and Method

The purpose of the study was to validate a self-assessment scale for teachers' evaluation. The scale was developed, discussed with experts and used to collect data for further validation process.

Instrument Development

Major objective of scale development is to compose a valid measure of an underlying construct (Clark & Watson, 1995)—teacher self-assessment in this study. Scale development process can be divided into three steps (Bearden, Netemeyer, & Sharma, 2003). In the first step, the construct is defined. In the second step, items are generated. In the third step, the construct validity is checked and revised if necessary (Daigneault & Jacob, 2014; Sousa & Rojjanasrirat, 2011).

Firstly, the teacher self-assessment construct was operationally defined. It is performance of teachers on five dimensions (standards) namely: subject matter knowledge, instructional planning and strategies, assessment, learning environment, and effective communication. Subject Matter Knowledge describes that teacher understands central concepts, principles, National Curriculum Standards and the methods of making subject matter accessible and meaningful for all students (Government-of-Pakistan, 2009b; Shulman, 1986; Stronge, 2010, 2013). Instructional planning and strategies cover teacher's planning of the content, selecting teaching materials, technology, and resources, engaging students in meaningful learning experiences (Government-of-Pakistan, 2009b; Stronge, 2010, 2013). Assessment provides diagnostic information regarding students' mental readiness for learning new content, provides formative and summative information needed to monitor student progress, helps keep student motivated, and recordkeeping purposes (Sanders, 2000). Learning environment includes climate of mutual trust and respect, motivation towards enhanced learning, minimum disruptions in teaching learning process (Danielson, 1996; Frey & Schmitt, 2007; Government-of-Pakistan, 2009b; Marzano et al., 1993). Communication is an ability to use appropriate language for teaching as per ability level

of students; communicate properly with students and parents, colleagues (Cornett-DeVito & Worley, 2005; Government-of-Pakistan, 2009b; Stronge & Tucker, 2003).

At the second stage, a 28 items based item pool was developed to reflect the aforesaid five standards of teachers' effectiveness. The scale was named *Self-assessment Instrument for Teacher Evaluation II*. A majority of the items was developed based on Stronge's (2010) work. Stronge used teacher quality indicators which are "tangible behaviours that can be observed or documented to determine the degree to which a teacher is fulfilling" the particular standard (p. 23). The researchers adapted such tangible behaviours and developed 33 items grouped into five dimensions. The response scales ranged from lowest to highest as 1= Never, 2= Rarely, 3= Sometimes, 4= Often, or 5= Always. It was assumed that the teachers who always practice tangible behaviors would be highly effective or vice versa.

Thirdly, content validity was sought through two panels. One expert panel comprised three professors of education who had more than 20 years of teaching experience in the field of teacher education and/or testing. The second panel comprised five practitioners—Secondary School Teachers (SSTs) of mathematics or English in a public high school in Pakistan—who had varying levels of teaching experience, ranging from 5 to 20 years. The expert panel determined if the items were clear and correctly grouped into the domains, or if the items were poorly worded or superfluous. The practitioners' panel determined whether they were able to understand the items clearly or had some questions regarding item clarity. Both the panels gave comprehensive feedback and opinion on the validity of the content, relevancy of the items to the certain domains, and redundancy between the items. In light of the critique sessions and feedback of the experts and practitioners, the SITE II was reduced to 28 items (see Table 1)

Table 1

Self-assessment Instrument for Teacher Evaluation-II

[Response Scale: Never (1) Rarely (2) Sometimes (3) Often (4) Always (5)]

I: Subject Matter Knowledge

- I demonstrate accurate knowledge of my subject matter
- I link content with past and future learning experiences
- I demonstrate a variety of skills of my subject area(s)
- I communicate content in ways that students can understand
- I use school and community resources to help students
- I teach according to the intellectual, emotional needs of the students
- I effectively address appropriate curriculum standards
- I base instruction on goals that reflect high expectations

II: Instructional Planning and Strategies

- I use strategies to enhance students' understanding
- I change teaching methodology to make topics relevant
- I understand individual differences of students and teach accordingly
- I use appropriate material, technology, and resources
- I engage, motivate, and maintain students' attention
- I teach the required curriculum according to time-table
- I use student learning data to guide planning

III: Assessment

- I conduct class tests to monitor student performance
- I evaluate students' performance and provide feedback
- I maintain students' results and use future improvement
- I revise content to enhance students' achievement
- I keep official record of students' learning progress

IV: Learning Environment

- I create a climate of mutual trust and respect in classroom
- I maintain a classroom setting that minimizes disruption
- I create friendly and supportive classroom environment
- I ensure students' participation in the learning process
- I encourage students to interact respectfully

V: Effective Communication

- I use correct vocabulary and grammar in speaking & writing
- I explain lessons according to the age and ability of students
- I respond to students' questions in appropriate language

Sample

The study sampled 279 secondary school teachers selected conveniently across 40 high schools in district Okara Pakistan. These teachers had taught English or mathematics to grade 9 and their students were now in grade 10 and waiting for their results in grade 9 to be announced in the Board of Intermediate and Secondary Education (BISE) Lahore. The response rate was 90%. Among respondents 119 (42.7%) were male while 160

(57.3%) were female. Years of teacher's experience ranged from 1 to 36 with the mean of 12.28. A vast majority of the teachers (86%) had a Master's Degree (academic degree) in some subject; only 13% had a Bachelor's Degree. Besides teachers data, the board roll numbers of 7245 students of these teachers were also collected from their teachers to get their achievement scores after their results were announced.

Data Collection

The data were collected in August and September 2014. One of the researchers (first author) personally visited 40 public high schools in district Okara—the schools which he could conveniently visit. The researcher met head teacher of each boys' high school, and received authorization from him or her to distribute the SITE-II among English or mathematics teachers in the school. After getting authorization from the head of the school, the research met the teachers and asked about their willingness to participate in this study. After the teacher showed interest in the project, the researcher distributed a consent letter to each teacher. The teacher read the consent form, put signature on that form, and returned it to the researcher. The researcher also gave a copy of that consent form to each teacher for the teacher's record. The first author then distributed the Self-assessment Instrument for Teacher Evaluation (SITE-II) to each teacher. After the teacher had completed the SITE-II, the data were collected and used for analysis purpose. Following this procedure the first author collected SITE-II from 160 male teachers. In the female schools, the researcher took the help of head teachers for data collection from 119 female teachers. Among 279 teachers, 137 were English teachers while 142 were mathematics teachers.

Data Analysis

An evidence of factor analysis and loadings can be given in the form of scree plot. The scree plot graphs the eigenvalues against the factor number. A scree plot displays the eigenvalues (greater than 1) associated with a component or factor in descending order versus the number of the component or factor. A factor analysis was conducted on 28 items. The scree plot showed that 5 of those factors explained most of the variability while the remaining factors explained a very small proportion of the variability.

Exploratory factor analysis was thought to be an adequate approach to understand the factors and loadings. Initially the researchers computed correlations of all pair-wise combinations of 28 items and determined that the results matrix of correlations was appropriate and of good quality for factor analysis by means of Bartlett's test of sphericity, $\lambda = 3823.247$, $df = 378$, $p = .000$, and a Kaiser–Meyer–Olkin measure of sampling adequacy, $KMO = 0.94$.

Principal component analysis was performed using the varimax rotation method for factor extraction on the items. A principal component analysis uses eigenvalues, which represent the proportion of variance accounted for by the factors. The eigenvalues greater than 1 showed that there were 5 factors that represented 59.25 % of the variance which is considered good. Subject matter knowledge explained 17.12 % of the observed variance, Instructional planning and strategies 13.78 %, Assessment 10.00%, Environment 9.19 %, and Effective communication explained 9.15% of the observed variance in Teacher evaluation construct. 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).

Correlations among five factors were calculated using the Pearson correlation (r). Table 2 shows that all the five factors significantly correlated with each other. The highest correlation was found between Subject matter knowledge and Instructional planning factors ($r = .76$), while least positive significant correlations was found between Subject matter knowledge and effective communication ($r = .44$). Summary of the results is given in Table 2.

Table 2

Correlations among Factors

Factors	1	2	3	4	5
Subject Matter Knowledge					
Instructional Planning and Strategies	.761**				
Assessment	.744**	.719**			
Learning Environment	.567**	.529**	.525**		
Effective Communication	.444**	.492**	.469**	.615**	

**Correlation is significant at the 0.01 level (2 tailed).

Confirmatory factor analysis was conducted using Lisrel and it was conducted in two stages as described by Byrne (2001), and Schumacker and Lomax (2004). Initially, confirmatory factor analysis was conducted in two stages. In the first step, the factors in the model corresponded to the items that were assumed to represent the components of the construct—teacher self-evaluation. In the second step, we examined the relationship between teacher self-assessment score and student achievement in English and mathematics. Chi-square index reported better fit, $\chi^2=643.14$, $p=0.0$. Since with chi-square test of model fit, the researcher may fail to reject an appropriate model in small sample size and reject appropriate model in large sample size (Gatignon, 2010), other measures are used for model fitness.

Other indexes used for fitness of the model comprise Root Mean Square Residual (SRMR)

which represents the average value, ranging from 0 to 1, across all standardized residuals, Goodness of Fit Index (GFI) which ranges from 0 to 1 and estimates the proportion of variability in the sample covariance matrix explained by the model, Comparative Fit Index (CFI), ranges from 0 to 1 and compares the model with the standard ‘null’ model that assumes zero population covariance among the observed variables, and Root-Mean-Square Error of Approximation (RMSEA), a value of less than 0.10 indicates a good model fit, which assesses a lack of fit of the population data to the estimated model. 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 (see figure 1).

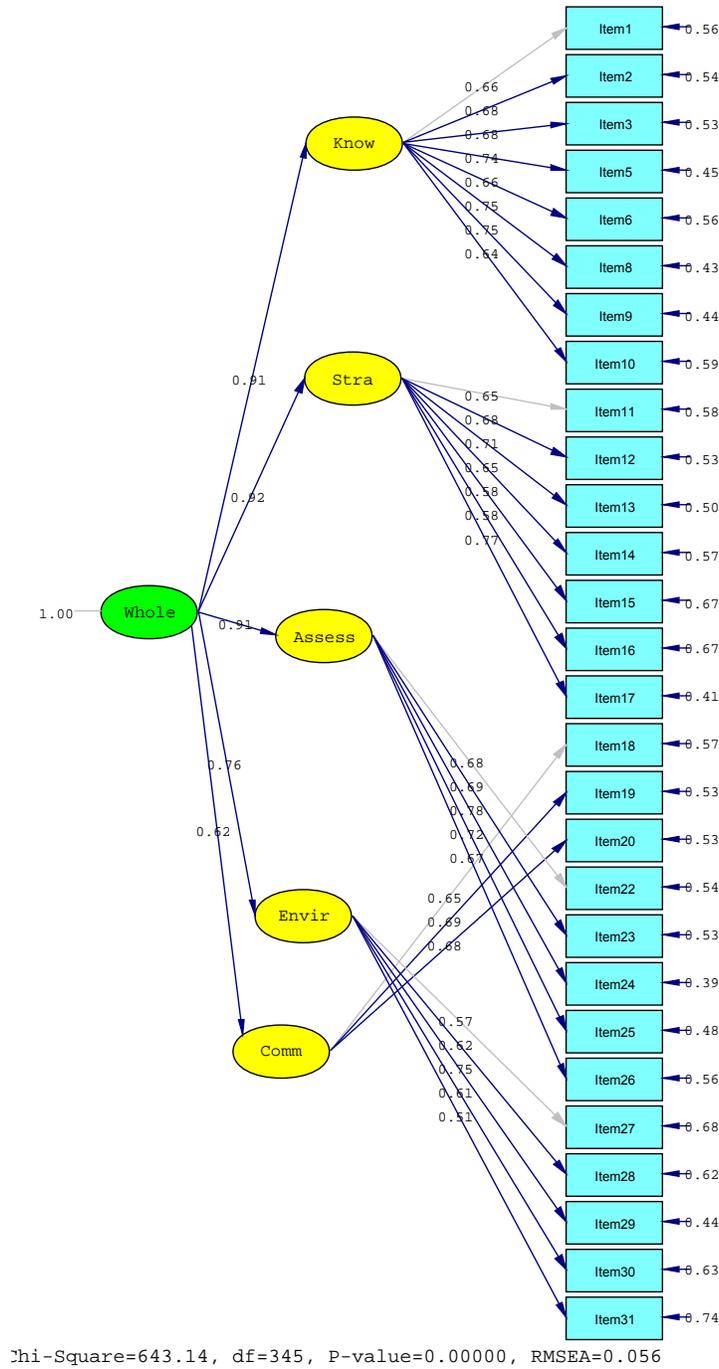


Figure 1: Confirmatory factor analysis: Standardized factor loadings and correlations. **Know**, Subject matter knowledge; **Stra**, Instructional planning and strategies; **Assess**, assessment; **Envir**, Learning environment; **Comm**, Effective communication. (n= 279).

The factor loadings, often referred to as validity coefficient, are estimated correlations which indicate how well a given item measures its corresponding factor. The loadings ranged from high to moderate—the lowest was 0.45, exceeding the factor-loading criterion of 0.35 described by Tabachnick and Fidell (2000).

At the second stage, the relationship was found between teacher self-assessment score and their students' scores in English and Mathematics obtained in the Board of Intermediate and Secondary Education Lahore Annual Examination 2014. Teachers' self-assessment scores on SITE II were significantly correlated with their student achievement in English (r ranged from .14 to .49) as well as in mathematics (r ranged from .13 to .34). The significant positive correlations provided criterion-related validity evidence.

Conclusions and Implications for Practice

Teacher self-assessment is a process in which teachers make judgments about the adequacy and effectiveness of their own knowledge, and performance. Self-assessment is used for formative evaluation, not summative, of teachers aiming at examining, altering, or improving practice (Kremer-Hayon, 1993). Teacher self-assessment is important as it extends assistance to new teachers, enhances career opportunities for veterans, and makes teachers more responsible for demonstrating their own competence. Self-assessment scales can measure teaching performance if developed diligently (Conceicao, Strachota, & Schmidt, 2007; Klecker, 2005). The process used in the development of the SITE-II is an example of diligence. *Self-assessment Instrument for Teacher Evaluation* is valid and reliable ($r = .94$) scale. The higher score on this scale represents higher effectiveness and lower score lower performance. Teachers can use this scale for evaluation of their own teaching and take remedial actions. SITE-II is an efficient tool that might be used as one of the data source of teacher evaluation. Mentors and supervisors can use it diagnostic purpose and designing professional development courses for teachers. However, it is suggested that self-assessment must be based on unbiased information in an environment of support and trust where teachers can take honest look at their practices. Directorate of Staff Development (DSD) might look at the worth of this questionnaire and

encourage head teachers to use it for teacher professional development purpose.

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95-135.
- Airasian, P. W., Gullickson, A. (2006). Teacher self-evaluation. In James H. Stronge (2nd Ed.), *Evaluating teaching: A guide to current thinking and best practice* (186-211). Corwin Press, Thousand Oaks, CA: Sage.
- Akram, M. (2012). *The relationship between teacher evaluation scores and student achievement in English and mathematics: Evidence from Pakistan*. (doctoral dissertation), University of Georgia.
- Almani, A. S. (2002). *Doctor of Philosophy in Education*. (Doctor of Philosophy), Hamdard University, Karachi.
- Aziz, M. A. (2010). *Effect of demographic factors and teachers competencies on the achievement of secondary school students in the punjab*. (Doctoral dissertation), Allama Iqbal Open University, Islamabad, Islamabad.
- Bearden, W. O., Netemeyer, R., & Sharma, S. (2003). *Scaling procedures: Issues and applications*: California: Sage.
- 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.
- Bibi, S. (2005). *Evaluation study of competencies of secondary school teachers in Punjab*. (doctoral dissertation), University of Arid Agriculture, Rawalpindi, Rawalpindi.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, 5(1), 7-74.
- Bolino, M. C., & Turnley, W. H. (2003). Counternormative impression management, likeability, and performance ratings: The use of intimidation in an organizational setting. *Journal of Organizational Behavior*, 24(2), 237-250.

- Boulter, P. (1987). The American experience: A report of a visit to study teacher appraisal and assessment in the United States in Autumn 1985. *School Organisation*, 7(2), 145-161. doi: 10.1080/0260136870070203
- Catt, S., Miller, D., & Schallenkamp, K. (2007). You Are the Key: Communicate for Learning Effectiveness. *Education*, 127(3), 369-377.
- Centra, J. A. (1973). Self-ratings of college teachers: a comparison with student ratings. *Journal of Educational Measurement*, 10(4), 287-295.
- Clandinin, D. J., & Connelly, F. M. (1988). Studying teachers' knowledge of classrooms: collaborative research, ethics, and the negotiation of narrative. *Journal of Educational Thought*, 22(2), 269-282.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, 7(3), 309-319.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119-142.
- Conceicao, S. C., Strachota, E., & Schmidt, S. W. (2007). The Development and validation of an instrument to evaluate online training materials. Retrieved from <http://files.eric.ed.gov/fulltext/ED504339.pdf>.
- Cornett-DeVito, M. M., & Worley, D. W. (2005). A front row seat: A phenomenological investigation of learning disabilities. *Communication Education*, 54(4), 312-333.
- Daigneault, P.-M., & Jacob, S. (2014). Unexpected but most welcome mixed methods for the validation and revision of the participatory evaluation measurement instrument. *Journal of Mixed Methods Research*, 8(1), 6-24.
- Danielson, C. (1996). *The framework for professional practice*. Alexandria, VA: : Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (2011) Evaluating teacher Evaluation: We Know About Value-Added Models and Other Methods. Phi Delta Kappan.
- 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.
- Della Fave, L. R. (1980). The meek shall not inherit the earth: Self-evaluation and the legitimacy of stratification. *American Sociological Review*, 45(6), 955-971.
- Dilshad, R. M. (2010). Assessing quality of teacher education: A student perspective. *Pakistan Journal of Social Sciences*, 30(1), 85-97.
- Ellett, C. D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101-128.
- Evertson, C. M. (1985). Training teachers in classroom management: An experimental study in secondary school classrooms. *The Journal of Educational Research*, 79(1), 51-58.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117-140.
- Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics*, 18(3), 402-423.
- Fullan, M. G. (1993). Why teachers must become change agents. *Educational Leadership*, 50(6), 12-12.
- 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.
- Gatignon, H. (2010). *Confirmatory Factor Analysis in Statistical analysis of management data*. DOI: 10.1007/978-1-4419-1270-1_4
- Ginsburg, M. (2009). *EQUIP2 State-of-the-Art Knowledge in Education*. Washington, DC: United States Agency for International Development.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.

- Government-of-Pakistan. (2009a). *National Education Policy 2009*. Islamabad Ministry of Education.
- Government-of-Pakistan. (2009b). *National professional standards for Teachers in Pakistan*. Islamabad: Policy and Planning Wing
- Guskey, T. R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6), 45-51.
- Guskey, T. R. (2007). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice*, 26(1), 19-27.
- Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types. *Journal of Research on Technology in Education*, 41(4), 393-416. doi: 10.1080/15391523.2009.10782536
- Heneman, R. L., Greenberger, D. B., & Anonyuo, C. (1989). Attributions and exchanges: The effects of interpersonal factors on the diagnosis of employee performance. *Academy of Management Journal*, 32(2), 466-476.
- 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.
- Hunt, G., Wiseman, D., & Touzel, T. J. (2009). *Effective teaching: preparation and implementation*. Spingfiled, Illinois. Charles C Thomas Publisher.
- Ingvarson, L. (2002). Development of a national standards framework for the teaching profession. Australian Council for Educational Research. Retrieved from http://research.acer.edu.au/cgi/viewcontent.cgi?article=1007&context=teaching_standards.
- Jumani, N. B. (2007). *Study on the competencies of the teachers trained through distance education in Pakistan*. (Post-Doctorate), Deakin University Australia Australia.
- Kauchak, D., Peterson, K., & Driscoll, A. (1985). An interview study of teachers' attitudes toward teacher evaluation practices. *Journal of Research & Development in Education*.
- Khan, B., Begum, S. (2012). Portfolio: A professional development and learning tool for teachers. *International Journal of Social Science and Education*, 2(2), 363-377.
- 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.
- Klecker, B. (2005). Assessing learning online: the top ten list. In C. Crawford et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2005* (pp. 119-121). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).
- Korthagen, F. A. (2004). In search of the essence of a good teacher: Towards a more holistic approach in teacher education. *Teaching and Teacher Education*, 20(1), 77-97.
- Kremer-Hayon, L. (1993). *Teacher self-evaluation: Teachers in their own mirrors*. Boston: Kluwer Academic.
- Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494-509.
- Laking, R. (2007). International Civil Service Reform: Lessons for the Punjab? Retrieved from http://prmp.pitb.gov.pk/sites/prmp.pitb.gov.pk/files/CSR_By_Rob_Laking.pdf.
- Lister, S., Bano, M., Carr-Hill, R., & MacAuslan, I. (2010). Country Case Study: Pakistan.
- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, 15(2), 129-151.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, 34(2), 120-131.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria: Association for Supervision and Curriculum Development.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model*. Alexandria, Virginia: ERIC.

McGreal, T. L. (1983). *Successful Teacher Evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.

Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *The Journal of Educational Research*, 80(4), 242-247.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33-53.

Milanowski, A. T., & Heneman, H. G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: a pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193-212.

Ngoma, S. (2011). Improving teacher effectiveness: an examination of a pay for performance plan for boosting student academic achievement in Charlotte-Mecklenburg Schools. Retrieved from <http://files.eric.ed.gov/fulltext/ED518815.pdf>.

Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin*, 88(639), 60- 79.

Peterson, K. D. (2006). Using multiple data sources in teacher evaluation systems. In H. J. Stronge (2nd Ed.), *Evaluating Teaching: A guide to current thinking and best practice* (pp. 212-230). Thousand Oaks, CA: Corwin.

Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices*. London: Corwin Press.

Peterson, P. L., & Comeaux, M. A. (1990). Evaluating the systems: Teachers' perspectives on teacher evaluation. *Educational Evaluation and Policy Analysis*, 12(1), 3-24.

Petty, G. (2009). *Teaching today: A practical guide* (4th ed.). Cheltenham: Nelson Thornes.

Quirk, T. J., Witten, B. J., & Weinberg, S. F. (1973). Review of studies of the concurrent and predictive validity of the National Teacher Examinations. *Review of Educational Research*, 89-113.

Sanders, W. L. (2000). Value-added assessment from student achievement data: opportunities and hurdles. *Journal of personnel evaluation in education*, 14(4), 329-339.

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57-67.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, 15(2), 4-14.

Sousa, V. D., & Rojjanasirrat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of evaluation in clinical practice*, 17(2), 268-274.

Stronge, J. H. (2006). Teacher evaluation and school improvement. In H. J. Stronge (2nd Ed.), *Evaluating Teaching: A guide to current thinking and best practice* (pp. 1-22). Thousand Oaks, CA: Corwin.

Stronge, J. H. (2007). *Qualities of effective teachers*. Alexandria, VA: : Association for Supervision and Curriculum Development.

Stronge, J. H. (2010). *Effective teachers= student achievement*. New York, NY: *Eye on Education*.

Stronge, J. H. (2013). *Teacher Performance Evaluation Program Handbook* Virginia: Fairfax County Public Schools.

Stronge, J. H., & Tucker, P. D. (2000). *Teacher Evaluation and Student Achievement. Student Assessment Series*. Texas: ERIC.

Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. New York: Eye on Education Larchmont, NY.

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of personnel evaluation in education*, 20(3-4), 165-184.

Tabachnick, B. G., & Fidell, L. (2000). *Using multivariate statistics* (4th edn.). Boston: Allyn and Bacon.

Tomlinson, C. A. (1999). Mapping a route toward differentiated instruction. *Educational Leadership*, 57(1), 12-17.

Varma, A., & Stroh, L. K. (2001). The impact of same-sex LMX dyads on performance evaluations. *Human Resource Management*, 40(4), 309-320.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership*, 41(8), 19-27.

Wang, K. H., Wang, T., Wang, W.-L., & Huang, S. (2006). Learning styles and formative assessment strategy: enhancing student achievement in Web-based learning. *Journal of Computer Assisted Learning*, 22(3), 207-217.

Wolf, K. (2006). Portfolios in teacher evaluation. In H. J. Stronge (2nd Ed.), *Evaluating Teaching: A guide to current thinking and best practice* (pp. 168-184). Thousand Oaks, CA: Corwin.

Zepeda, S. J. (2006). Classroom-based assessments of teaching and learning. In H. J. Stronge (2nd Ed.), *Evaluating Teaching: A guide to current thinking and best practice* (pp. 101-124). Thousand Oaks, CA: Corwin.

Zepeda, S. J. (2014). *The principal as instructional leader: A handbook for supervisors*. New York: Routledge.