

Teacher evaluation by different internal evaluators: Head of departments, teachers themselves, peers and students

Talal S. Almutairi¹, Nawaf S. Shraid²

¹Ph.D. in Educational Evaluation, Assessment and Effectiveness, Kuwait

²Ph.D. in Educational Foundations and Head Department in Middle School, MOE Kuwait, Kuwait

Article Info

Article history:

Received Aug 8, 2020

Revised Feb 28, 2021

Accepted Apr 13, 2021

Keywords:

Evaluator
Peer-evaluation
Quality teaching
Self-evaluation
Student evaluation
Teacher performance

ABSTRACT

This study analyzed teacher evaluation in school, through involving different internal evaluators, in order to determine the extent to which they evaluate teacher performance accurately and objectively. Evaluation survey instruments are used in this study, which are designed based the criteria of existing teacher evaluation system in the context, along with other criteria for evaluating teachers. The sample of this study included teachers, heads of departments and students from high schools in four different districts in Kuwait, received responses as 100 from heads of department, 100 from teachers 'self-evaluation', 100 from peer and 912 from students. The findings show that there is no significant difference between teachers' self-evaluation and heads of departments' evaluation. On the other hand, this study finds that subjectivism and competition may have an effect on peer evaluation and students may over-evaluate their teachers' performance as attempt to draw a better picture of their teachers in front of evaluators.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Talal S. Almutairi
Educational Evaluation, Assessment and Effectiveness
Kuwait
Email: dr.talmutairi@gmail.com

1. INTRODUCTION

The phrase 'teacher evaluation' is used to refer to what teachers have achieved and what they need to develop or improve in their performance. We also define it as an image that helps the educationists or policymakers or inspectors or others stakeholders to find out about teachers and their teaching. Nolan and Hoover [1] define teacher evaluation as having a designed function to make decisions about teachers' performance clear for either development or sanctions or rewards. Therefore, teacher evaluation should be used effectively to meet its aims or purposes. Almutairi [2] sets out several principles that make teacher evaluation well organized, functional and effective: maintaining a balance between formative and summative purposes, clarifying the criteria of teacher evaluation and teachers' performance should be reflected by using several tools and involving internal/external evaluators.

This study builds on a research study by Almutairi, Tymms and Kind [3] that focused on identifying what tools of teacher evaluation should be used. Their findings, based on teacher perspectives, were that observation by heads and inspectors, students' achievement, teachers' portfolio, self, peer and student evaluation either survey or by interview should be used to evaluate teachers' performance. This study analyses the data collected from internal evaluators (head department, teacher themselves, peer, and students) in order to determine the extent to which they evaluate teacher performance accurately. The study addresses the following questions: 1) To what extent does each internal evaluator or group of evaluators evaluate

teacher performance accurately? 2) Are there any potential difficulties to or drawbacks of involving any one of these internal evaluators?

With regard to the contribution of this study, firstly, given that teacher evaluation is an essential component in the educational field for both development and improvement targets, it makes a general contribution towards furthering this goal. This study is also intended to contribute towards making teacher evaluation systems more effective in identifying both outstanding and under-performing teachers. Moreover, as mentioned, this study takes the research conducted by Almutairi, *et al.* [3] one step further.

Teachers' performance or teaching effectiveness is seen as a key factor or challenge for development and improvement, especially in terms of developing a new vision of an education system. To evaluate teachers' performance, decision-makers need to collect data that can determine what a teacher has done or achieved during the school year. There are a lot of potential data sources that are used to determine teacher performance such as student evaluation or rating, peer and self- evaluation, and administrator/heads rating [2], [4]. The benefit of using multiple sources rather than a single source of data, is that evaluators or decision-makers can be provided with a picture of teacher practice that is closer to the realities of day-to-day teaching practices, based upon which appropriate and effective decisions can be made. As Zhang [5] argues, using multiple sources can validate one another, thereby reducing bias and increasing validity. Reliability can also be increased by using multiple sources as each one has a different margin of error [6]. Furthermore, using multiple sources reflects the multifaceted nature of teacher performance and as such, can also be useful to teachers in the form of feedback that identifies their weaknesses and strengths [7]. This may explain why teachers in the study [3] support the use of multiple tools of evaluation in their performance evaluation. Teachers in the study primarily favored classroom observation by heads or inspectors as a tool of performance evaluation, as well as students' achievement, self-evaluation, peer evaluation and teaching portfolio. While there was less support for student evaluation compared to these other tools, teachers did favor the idea of allowing their students to provide opinions about their teaching.

It appears that in selecting appropriate evaluations, the level of education is a consideration. For example, Barnett, Matthews, and Jackson [8] stated comparison between student evaluation and self-evaluation of teachers in higher education, 12 evaluation items were given to students and 31 items regarding teaching ability, teaching materials, content organization and teacher knowledge, were given to the teachers; no significant differences were found between the results of student evaluation and teacher self-evaluation results. By contrast, in a study comparing student evaluation and self-evaluation and heads of departments' evaluation in high schools, Eid [9] found some significant differences. The mean of student evaluation was higher than the mean of evaluation by teachers themselves and heads of departments, which suggests that students at this level of education see their teachers better than teachers see themselves or heads of department. Eid claimed that students may also be deliberately presenting a positive picture of their teacher in front of the evaluators. This hypothesis is indicated in previous studies which found that younger students tend to evaluate their teachers-based personality rather than practice, as lower scores were given to strict teachers and higher scores to teachers who were perceived as considerate of them [10]. Elsewhere, Peterson, Wahlquist and Bone [11] indicate that students are able to evaluate teachers based on their teaching but that this requires the use of items of evaluation that are appropriate for the student level.

In terms of peer evaluation, as found by previous studies, peer feedback is one of several ways in which teachers can share their experience and knowledge with each other; feedback from peers can be helpful to change or develop their teaching, and exchanging written (both for the teacher writing it and the teacher receiving it) can be a useful professional development exercise [2], [12], [13]. Peer evaluation can also be utilized in a summative way by reflecting how teaching is delivered by their school staff to students and involving appraisal or personal decisions [14], [15]. However, using peer evaluation in this way can be problematic for teachers as it requires high levels of trust [16]. In addition, Zhang [5] point out that personal relationship can influence evaluation when used to make judgment or decision about colleagues; bias and subjectivity can be increased when there is misunderstanding among teachers, thereby further undermining the value of peer evaluation [12]. Thus, from the teachers' perspectives, peer evaluation seems to be more valuable when used formatively [3]. Nonetheless, if used for summative or formative purposes and to mitigate against bias and subjective in peer evaluation, consistency of interpretation will be strengthened by clear performance guidelines and explicit criteria of teaching [12], [17].

With regard to the role of heads of departments, Almutairi [2] found that they are more involved in teacher evaluation compared to head teachers and inspectors and that teacher valued their evaluative role far more highly than other internal or external evaluators. One of the reasons for this, Almutairi explains, is that heads of departments are more specialized. One head teacher interviewed in his research acknowledged that a lack of subject knowledge was a problem in terms of evaluating teachers in their various subject specialisms. For example, originally a teacher of Arabic, he found himself being expected to evaluate teachers of English and did not feel qualified to do so. The validity of evaluations carried out by heads of department and

teachers 'self' is supported by a study that compared evaluation by heads of department evaluation and self-evaluation which found no significant differences between the reports [9]. In contrast, other studies argue that self-evaluation provides an inaccurate reflection of teachers' performance compared to other forms of evaluation, due to the possibility of misrepresentation and misreporting [18] and because teachers' view of their own performance can vary from day to day [19]. Furthermore, Zhang [5] found that some teachers may over-rate themselves – in their view – they just think, they deserve a high score.

2. RESEARCH METHOD

This research is statistically descriptive in nature because it is grounded in the need to describe and interpret and is concerned with conditions or relationships that exist [20]. Furthermore, it is used to determine differences and compare similarities [21].

2.1. Instrument

Before constructing the instruments for this study, the researcher reviewed the relevant literatures pertaining to teacher evaluation. Current teacher evaluation criteria in Kuwait as described in Teacher Evaluation System No.36/2006 [22] were reviewed, along with other criteria for evaluating teachers such as [1], [23]-[25]. However, the actual evaluation survey instruments used for this study are unique.

Closed questions were chosen to build the survey instrument as this type of question makes it easier to draw comparisons between groups, as well as being easy to complete by participants [21], [26]. Four survey instruments were designed for teachers, peers, heads departments, and students respectively, in order to make these appropriate for each participant group. Participants were asked about the following aspects: instruction, assessment, management, students, and interaction & cooperation. Each sub-section comprised five items, except for the last category (interaction and cooperation) which comprised four items, totaling 24 items. The items were measured on a 5-point Likert-type scale (from never to always/from far below to far above standards).

The instruments were designed in English and were then translated from English to Arabic by the researcher, then back translated by three professional translators to ensure that the translations were accurate and were appropriate for a context in which the official language is Arabic. No significant discrepancies were founded or reported by translators.

2.2. Sample and procedure

The target population for this study was head of department, teachers, peer and students in high school. Sixteen high schools from four districts in Kuwait (4 of 6 districts) were randomly chosen as: four high schools from each district, of which two schools were taught by female teachers and two schools taught by male teachers.

The researcher obtained consent and permission to collect participants' responses during the second semester of the 2018–2019 academic year from the Ministry of Education in Kuwait (MOE). The evaluation survey instruments were distributed by visiting schools and asking the school administration to meet different heads of departments in order to evaluate two teachers' performance in their departments and then ask those two teachers who had been evaluated by their heads of department to participate in self and peer evaluation. After that, the researcher asked the school administration to hand out the instrument to about 10 students in classes of those teachers who participated in self and peer evaluation, to evaluate their teacher's performance. The researcher obtained 300 responses from teaching staff as: 100 from heads of departments, 100 from teachers 'self-evaluation', 100 from peers, and 912 from students.

2.3. Validity

According to Fraenkel, Wallen and Hyun [27], validity is "the appropriateness, correctness, meaningfulness and usefulness of the specific inferences researchers make based on the data they collect". Thorndike and Thorndike-Christ [28] explain different types of validity exist in educational research such as content, construct and criterion. In this research, content was tested through collecting evidence by experts in the area who served as judges, as this form of validity is useful when the questions are well known and identifiable [29]. The items of the evaluation instruments were discussed with four academic staff in the School of Education at Kuwait University and five heads teacher in the Kuwaiti MOE. They were asked whether the items were well presented and clear, whether they addressed the study aims, and whether they could generate suitable answers. The survey instruments were judged to be appropriate, correct and meaningful for the teachers as self and peer evaluators and for the heads of department and students.

A pilot study with three teachers, three peer teachers, three heads of department and four students was also carried out: participants were interviewed after completing the evaluation instruments in order to

validate the construct in terms of how participants interpreted the criteria and how well instrument worked, as recommended in [29]. Participants were asked to report their experience of filling in the survey and what were they thinking as they did so. Their responses in these interviews confirmed that the instruments were measuring and reflecting accurate responses.

2.4. Reliability

Cronbach's alpha was used to shed light on the results of the internal consistency reliability of aggregated scales. According to Cohen, Manion and Morrison [21], an alpha coefficient greater than .90 indicates that the instrument is very highly reliable, while an alpha of less than .70 indicates an unreliable instrument. Table 1 shows that the reliability of the peer measure regarding interaction and cooperation was .79, which indicates that this scale was reliable. Other scales measuring were found to be between very highly reliable and highly reliable. Test-retest was also used to measure the reliability of the instrument in this study. For this test-retest, the sample comprised five teachers, five heads of departments, five peer and five students. The test-retest period was 12 days, as recommended by Cooper and Schindler [30] who suggest that the period between the first and the second application should be not too short or too long.

A t-test was used to analyze test-retest results, as indicated in [21]. Table 2 shows the result of the paired t-test, determining the mean responses for the aggregated scales across two periods. The p-value was set at >.05 for all scales, indicating no statistical difference between the participants' responses in the scores on the test and retest. Therefore, the survey instruments were reliable.

Table 1. Cronbach's alpha results for the aggregated scales

Scales	Number of items	Sample	Cronbach's alpha
Instruction	5	Head of department	.91
		Teacher	.90
		Peer	.87
		Student	.93
Assessment	5	Head of department	.91
		Teacher	.89
		Peer	.84
		Student	.92
Management	5	Head of department	.89
		Teacher	.91
		Peer	.82
		Student	.92
Students	5	Head of department	.91
		Teacher	.90
		Peer	.90
		Student	.94
Interaction & cooperation	4	Head of department	.90
		Teacher	.87
		Peer	.79
		Student	.94

Table 2. Paired Sample test for test-retest

Scales	Mean difference (1st & 2nd)	Std. deviation	Std. error mean	95% interval of the difference		t	DF	p-value
				Lower	Upper			
Instruction	.12000	.89772	.20074	-.30014	.54014	.598	19	.557
Assessment	-.18000	.82819	.18519	-.56760	.20760	-.972	19	.343
Management	-.03000	.29218	.06533	-.16674	.10674	-.459	19	.651
Students	.16000	.65083	.14553	-.14460	.46460	1.099	19	.285
Interaction & cooperation	.12500	.68104	.15229	-.19374	.44374	.821	19	.422

3. RESULTS AND DISCUSSION

3.1. Results

The results of the inferential analysis, Table 3 shows a statistically significant difference between internal evaluators with regards to instruction, assessment, management, student, interaction and cooperation. Each aggregated scale has a p-value of <.05. Therefore, Tukey's HSD post hoc test was chosen to calculate the mean difference between heads of departments' evaluation of the teacher, self-evaluation, peer evaluation and students' evaluation.

Table 3. ANOVA results of sample groups (teachers, heads of departments, peer and students)

Scale	Grouping variable	Sum of squares	df	Mean square	F	P-value
Instruction	Between groups	167.266	3	55.755	33.096	.000
	Within groups	2035.088	1208	1.685		
	Total	2202.354	1211			
Assessment	Between groups	77.227	3	25.742	14.410	.000
	Within groups	2158.025	1208	1.786		
	Total	2235.253	1211			
Management	Between groups	126.623	3	42.208	23.798	.000
	Within groups	2142.473	1208	1.774		
	Total	2269.096	1211			
Student	Between groups	74.782	3	24.927	12.482	.000
	Within groups	2412.460	1208	1.997		
	Total	2487.242	1211			
Interaction & cooperation	Between groups	213.437	3	71.146	41.312	.000
	Within groups	2080.371	1208	1.722		
	Total	2293.808	1211			

Table 4 shows the mean difference between internal evaluators with regards to instruction as an aggregated scale. Two opposite patterns are revealed: the distribution of student evaluations is skewed at the higher end as the mean value is ($m=3.66$) while the distribution of peer internal evaluations is skewed towards the lower end of the scale as the mean value is ($m=2.31$). Other evaluators fall somewhere in between the students and peers with regards to instruction.

This data suggests that students tend towards providing a better picture of teacher performance than teachers themselves in terms of instruction. As each evaluator was asked about teacher performance regarding to understanding the central concepts, representing content accurately for students, linking content with student experiences, achieving learning targets, planning lessons in a justifiable sequence. Otherwise, peer evaluators have another point view about their colleague performance that is different about what students reflect, what head of department and teacher themselves reflect about teachers' performance in term of instruction.

Table 4. Tukey's HSD post hoc test, instruction aggregated scale

Instruction	N	Mean	Mean difference		
			Teacher	Peer	Student
Head department	100	3.32	.058	.980*	.341
Teacher	100	3.26		.922*	.399*
Peer	100	2.34			1.32*
Student	912	3.66			

In term of assessing students in the classroom, internal evaluators are asked to evaluate teacher performance with regard to the following criteria: using assessment tools that are congruent with curriculum targets; assessing student understanding during instruction; providing feedback to students in a timely manner; using assessment data in order to reteach students if needed; recording student achievement and handling records. Table 5 shows a positive correlation between self-teacher evaluation, head of department and student, as there are no significant differences between the rating from these evaluators. By contrast, there is a significant difference between peer reports and other evaluators' reports regarding assessment, suggesting that peers perceive their colleagues less favourably in terms of their assessment practices when compared to teachers' self- perceptions and the perceptions of other internal evaluators.

Table 5. Tukey's HSD post hoc test, assessment aggregated scale

Assessment	N	Mean	Mean difference		
			Teacher	Peer	Student
Head department	100	3.44	.046	.720*	.189
Teacher	100	3.39		.674*	.235
Peer	100	2.72			.909*
Student	912	3.62			

Table 6 presents the evaluations of the teachers' ability to manage their classrooms, from the different internal perspectives. The results show a difference between peer evaluators and others in this item, which includes the following criteria: the teacher's ability to create a classroom environment that is respectful for students; the teacher's ability to organize the classroom environment to support learning activities; the teacher's ability to engage and encourage students in productive learning activities; the extent to which the teacher manages and maximizes lesson time; the teacher's ability to deal with and control student behavior. As with the previous item, peer evaluators judge the performance of their colleagues more harshly than they judge their own and how their performance is rated by heads of departments and their students.

Table 6. Tukey's HSD post hoc test, management aggregated scale

Management	N	Mean	Mean difference		
			Teacher	Peer	Student
Head department	100	3.28	.056	.760*	.379*
Teacher	100	3.33		.816*	.323
Peer	100	2.52			1.13*
Student	912	3.65			

In term of dealing with students, Table 7 shows differences between teacher self-evaluations and their students, who rate their teacher's performance far higher, and differences between teacher self-evaluation and peer evaluation in which colleagues rate each other's performance far lower. This item included the following criteria: asking and answering effective questions that promote in-depth understanding; helping students to recognize the value of subject matter; giving students the opportunity to think, to talk and some responsibilities inside the classroom; and communicating clearly and effectively with students. As in Table 4, the distribution of student evaluations is skewed towards the higher end of the scale while the distribution of peer evaluations is skewed towards the lower end of the scale.

Table 7. Tukey's HSD post hoc test, student aggregated scale

Student	N	Mean	Mean difference		
			Teacher	Peer	Student
Head department	100	3.36	.078	.500	.324
Teacher	100	3.28		.422	.402*
Peer	100	2.86			.824*
Student	912	3.68			

Table 8 shows how teachers were rated in terms of interaction and cooperation in the classroom. This item includes: a teacher's interactions with colleague, parents and student; a teacher's understanding of and respect for the different beliefs, traditions and value of others; a teacher's participation in the school community; and contribution to non-duties teaching. The results show that teachers are overrated in this item by students compared to teacher's self-evaluation and the evaluation of heads of department, whereas the distribution of peer evaluations is again skewed toward the lower end of the scale.

Table 8. Tukey's HSD post hoc test, interaction and cooperation aggregated scale

Interaction & cooperation	N	Mean	Mean difference		
			Teacher	Peer	Student
Head department	100	3.20	.112	.925*	.502*
Teacher	100	3.08		.812*	.615*
Peer	100	2.27			1.42*
Student	912	3.70			

3.2. Discussion

There is much debate about how teacher evaluation can be effective in terms of improving individual decision-making, development and improvement of teaching. Involving different internal evaluators in this process is a way of making the system of teacher evaluation system more effective. Findings in this investigation are grouped into three areas for the purpose of this discussion: the accuracy of peer evaluation compared to self-evaluation and head of department evaluation; teacher's self-evaluation

compared to head of department evaluation; student evaluations compared to both teachers' self-evaluations and heads of departments' evaluations.

Firstly, peer evaluation in teacher performance evaluation should be used only for formative purposes because using it for summative purposes is problematic and potentially divisive. In this sense, this study supports [3], [16], [31] who conclude that teachers' peer evaluations should be used formatively and not be the basis for decisions about promotion or tenure. The tendency for peers to judge their colleagues harshly in this study suggests that the accuracy of peer evaluation can be undermined by subjectivity, preference or prejudice, and competition for promotion. Therefore, the findings in this study are in contrast with [17], [12], who suggested that explicit criteria and clear evaluation guidelines when evaluating each other might be helpful in avoiding bias and subjectivity. Similarly, the findings do not support [32], in discussing best peer evaluation practice, suggested informing teachers that they work with each other rather than against each other. In this study, the evaluation survey instrument consisted of explicit criteria and guidelines for peer evaluation and teachers were informed about the importance and benefits of peer evaluation when they work with each other. The findings in this study suggest that when teachers are asked to evaluate themselves and their colleagues, they do feel in competition with one another and will not rate their colleague's performance as well as they rate their own performance. Consequently, including peer evaluation in performance evaluation systems is not effective nor accurate when the purpose of the evaluation is summative.

Secondly, this study shows consistency between teacher self-evaluations and head of department evaluations, which is in line with the results of Eid's study [9]. However, its findings contrast previous findings that indicate self-evaluation may lead to inaccurate reflection or misreporting and over-rating by teachers [5], [18], [19] in that teachers in this study seemed able to evaluate their performance accurately without any exaggeration. This may have been because teachers knew that their self-evaluations would be compared with the evaluations of others, notably, their head of department. Heads of department are likely to be more involved in the teacher evaluation process than others evaluators, and they are seen as experts and specialized in their subject [2], so teachers are likely to respect their views. Previous literature has suggested that teachers' self-evaluations are susceptible to misreporting or misrepresentation and that there is too much variation in the level of feelings from day to day that undermines the validity of these evaluations and that this could be avoided by using video/audio recordings of lessons or writing down what happened during their lesson and using these as sources of data to evaluate themselves [33]. This investigation has shown that using explicit, well organized and clear criteria or checklists in the self-evaluation survey instrument can also help and support teachers to reflect accurately on their performance, thereby avoiding misrepresentations of their performance.

Third, when evaluators or policymakers involve student in giving opinions or views about their teacher performance, they should be cautious. As the results of this study show, students tend to over-evaluate their teachers when compared with teachers' self-evaluations or those of heads of departments in most scales of evaluation. In this regard, the findings of this study are consistent with research [9] which found that students want to show their teachers in a good light when under scrutiny from outside evaluators. Therefore, students' evaluations of teaching need to be interpreted with caution and should not be seen as primary sources that accurately reflect teachers' performance. In this sense, within the teacher evaluation system, students' evaluations should be not given equal weight to other tools and perspectives, but may be better used to identify patterns in teacher performance and compare to other evaluations such as self-evaluations or heads of departments reports when drawing conclusions about overall performance. It should also be taken into account that just as students can over-rate teachers they like or respect, they may also under-rate teachers they dislike. However, student evaluation could be fruitfully used for formative purposes to provide feedback to teachers, where evaluation is seen as for teachers rather than against them [34], [35], this reduces some of the concerns and anxieties teachers may have when evaluated by their students [35].

4. CONCLUSION

The research found that there is no significant difference between teachers' self-evaluation and heads of departments' evaluation. On the other hand, this study finds that subjectivism and competition may have an effect on peer evaluation and students may over-evaluate their teachers' performance as attempt to draw a better picture of their teachers in front of evaluators.

This study has given an account of teacher evaluation that involves various internal evaluators. It has argued that peer evaluation should be used for formative purposes only as using it for summative purposes leads to subjectively and is often overly influenced by the negative effect of competition. Furthermore, due to the tendency of over-rating teachers' performance, student evaluation should also be used for formative purposes, or to identify patterns and tendencies that can then be compared with other

evaluative tools. By contrast, teachers can and should be involved in evaluating their own performance, as this study has shown that they are able to do so quite accurately and in a way that aligns with the evaluations of their heads of department.

Finally, this research recommends that further research be undertaken as: 1) Conduct similar research in higher education institutions. As based on our knowledge, most universities and colleges in Kuwait or elsewhere only apply student evaluation and some of them also apply peer evaluation, we suggest that self-evaluation, peer evaluation, head of department evaluation and student evaluation be compared in higher education; 2) Compare data over several years and in different time periods of existing systems at school level that apply peer, self, student, heads of departments evaluations. Conduct more studies about combination of evaluation data sources, based on the perspectives of stakeholders (such as holistic, numerical, and portfolio or matrix approaches).

REFERENCES

- [1] J. Nolan and L. Hoover, *Teacher supervision and evaluation: Theory into practice*, 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc., 2009.
- [2] T. Almutairi, "Teacher Evaluation in Kuwait-Evaluation of the Current System and Consideration of Risk-Based Analysis as a Principle for Further Development," Doctoral Dissertation, Durham University, Durham, UK, 2016.
- [3] T. Almutairi, P. Tymms, and P. Kind, "The tools of teacher evaluation: What should be used in teacher evaluation from the teachers' perspective," in *International Business and Education Conferences Proceedings*, London, UK, 2015.
- [4] R. Berk, "Survey of 12 strategies to measure teaching effectiveness," *International Journal of Teaching and Learning in Higher Education*, vol. 17, no. 1, pp. 48-62, 2005.
- [5] X. Zhang, "The role of teacher appraisal in teacher professional development: A case study in schools in Shanghai," Doctoral Dissertation, The University of Hong Kong, Pokfulam, Hong Kong, 2008.
- [6] C. DePascale, "Managing multiple measures," *Principal*, vol. 91, no. 5, pp. 6-10, 2012.
- [7] A. Burnett, E. Cushing and L. Bivona, "Uses of multiple measures for performance-based compensation," Washington, DC: Center for Educator Compensation Reform, 2012.
- [8] C. Barnett, H. Matthews, and R. Jackson, "A Comparison between student ratings and faculty self-ratings of instructional effectiveness," *American Journal of Pharmaceutical Education*, vol. 67, no. 4, pp. 1-6, 2003.
- [9] G. Eid, "Teacher evaluation in high School in Kuwait: A comparative study of self-evaluation, student evaluation and heads of departments," *Educational Journal: Kuwait University*, vol. 19, no. 76, pp. 79-149, 2005.
- [10] S. Liu and C. Teddlie, "A follow-up study on teacher evaluation in China: Historical analysis and latest trends," *Journal of Personnel Evaluation in Education*, vol. 18, no. 4, pp. 253-272, 2005.
- [11] K. Peterson, C. Wahlquist and K. Bone, "Student surveys for school teacher evaluation," *Journal of Personnel Evaluation in Education*, vol. 14, no. 2, pp. 135-153, 2000.
- [12] A. Salih, "Peer evaluation of teaching or fear evaluation: In search of compatibility," *Higher Education Studies*, vol. 3, no. 2, pp. 102-114, 2013.
- [13] R. Eri, "Peer observation of teaching: Reflections of an early career academic," *Universal Journal of Educational Research*, vol. 2, no. 9, pp. 625-631, 2014.
- [14] C. Fernandez and J. Yu, "Peer review of teaching," *Journal of Chiropractic Education*, vol. 21, no. 2, pp. 154-161, 2007.
- [15] J. Blackmore, "A Critical evaluation of peer review via teaching observation within higher education," *International Journal of Educational Management*, vol. 19, no. 3, pp. 218-232, 2005.
- [16] M. Joshua, A. Joshua, B. Basse, and I. Akubuiro, "Attitude of Nigerian secondary school teachers to peer evaluation of teachers," *Teacher Development: An International Journal of Teachers' Professional Development*, vol. 10, no. 3, pp. 331-341, 2006.
- [17] S. Johnson and S. Fiarman, "The potential of peer review," *Educational Leadership*, vol. 70, no. 3, pp. 20-25, 2012.
- [18] L. Goe, C. Bell and O. Little, "Approaches to evaluating teacher effectiveness: A research synthesis," Washington, DC: National Comprehensive Centre for Teacher Quality, 2008.
- [19] K. Looft, "Teachers' perceptions of the evaluation process," Doctoral Dissertation, University of Georgia, Athens, US, 2002.
- [20] J. Best and J. Kahn, *Research in education*, 10th ed. Boston: Pearson Education, 2005.
- [21] L. Cohen, L. Manion and K. Morrison, *Research method in education*, 6th ed. London: Routledge, 2007.
- [22] Kuwait Teachers Society, "Guidance and the policy of teacher evaluation: Number 36/2006," *Kuwait Teacher Society Journal*, vol. June, no. 1586, pp. 1-15, 2010.
- [23] E. Iwanicki, "Evaluation in supervision," in G. Firth and E. Pajak, Eds., *The handbook of research on school supervision*. New York, Macmillan Reference Library, 1998.
- [24] C. Danielson, *Enhancing professional practice: A framework for teaching*. Alexandria, VA: ASCD, 2007.
- [25] W. Ribas, *Teacher evaluation that works: The educational, legal, public relations [political] & social-emotional [E.L.P.S] standards & processes of effective supervision & evaluation*, 2th ed. Westwood, MA: Ribas Publications, 2005.
- [26] P. Tymms, "Questionnaire," in R. Coe, M. Waring, L. Hedges, and J. Arthur, Eds., *Research methods and methodologies in education*. London, SAGE Publication, 2012.

-
- [27] J. Fraenkel, N. Wallen, and H. Hyun, *How to design and evaluate research in education*, 8th ed. New York: McGraw-Hill, 2011.
- [28] R. Thorndike and T. Thorndike-Christ, *Measurement and evaluation in psychology and education*, 8th ed. London: Pearson Education, 2009.
- [29] J. Creswell, *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*, 4th ed. Boston: Pearson Education, 2012.
- [30] D. Cooper and P. Schindler, *Business research methods*, 7th ed. New York: McGraw, 2000.
- [31] K. Schultz and D. Latif, "The planning and implementation of a faculty peer review teaching project," *American Journal of Pharmaceutical Education*, vol. 70, no. 2, pp. 1-6, 2006.
- [32] S. Golparian, J. Chan, and A. Cassidy, "Peer review of teaching: Sharing best practices," *Collected Essays on Learning and Teaching*, vol. 8, pp. 211-218, 2015, doi: 10.22329/celt.v8i0.4239.
- [33] V. Nikolic, "Self-evaluation and improved teaching practice," USA: ERIC Document Reproduction Services, 2002.
- [34] C. Mertler, "Students as stakeholders in Teacher evaluation; teacher perceptions of a formative feedback model," Paper presented at *Annual Meeting of the Mid-Western Educational Research Association*, Chicago, USA, 1997.
- [35] B. Burr, "Student Voices in Teacher Evaluations," Doctoral Dissertations, Brigham Young University, Provo, Utah, 2015.