Teacher Evaluation Training and Its Effect on EFL Teacher Evaluators' Perceptions and Practices[1]

Mohsen Shafaghi[2] & Masoomeh Estaji[3], Allameh Tabataba'i University, Tehran, Iran

Abstract

The current study sought to examine whether explicit instruction affects teacher evaluators in terms of a six-factor Teacher Evaluation (TE) model encompassing Perception, Method, System, Content, Purpose, and Outcome. Moreover, it explored how training of the evaluators affects their practices in actual evaluation. To this end, twenty EFL teacher-evaluators selected as the participants of the study filled out a TE questionnaire (developed by the researchers) and participated in a semi-structured interview. In addition, the evaluators' actual practice was examined through recording of post-observation debriefing sessions between the evaluators and the teachers they supervised. Subsequently, a four-session training course on various aspects of teacher evaluation was held. After four weeks, the interviews, questionnaires, and debriefing sessions were repeated. To analyze the data, a two-way repeated measures Analysis of Variance (ANOVA) was run and the comparison of the results showed that the training course had positively affected many aspects of evaluators' perceptions and practices. Likewise, the content analysis of the interviews and debriefing sessions identified the features of the training course that the evaluators found most valuable such as the methods of observation, the strategies required for overcoming the evaluators' communication problems with teachers and managing the debriefing sessions. The results also revealed that the TE factor outcome was not affected by the training instruction course.

Resumen

El presente estudio buscó examinar si la instrucción explícita afecta a los evaluadores de maestros en términos de un modelo de Evaluación de Maestros (TE) de seis factores que abarca Percepción, Método, Sistema, Contenido, Propósito y Resultado. Además, exploró cómo la formación de los evaluadores afecta sus prácticas en la evaluación real. Para ello, veinte profesores-evaluadores de inglés como lengua extranjera seleccionados como participantes del estudio llenaron un cuestionario TE (desarrollado por los investigadores) y participaron en una entrevista semiestructurada. Además, se examinó la práctica real de los evaluadores a través de la grabación de sesiones informativas posteriores a la observación entre los evaluadores y los profesores que supervisaban. Posteriormente, se llevó a cabo un curso de formación de cuatro sesiones sobre diversos aspectos de la evaluación docente. Después de cuatro semanas, se repitieron las entrevistas, los cuestionarios y las sesiones informativas. Para analizar los datos, se ejecutó un Análisis de Varianza (ANOVA) de medidas repetidas de dos vías y la comparación de los resultados mostró que el curso de capacitación había tenido un impacto positivo sobre muchos aspectos de las percepciones y prácticas de los evaluadores. Asimismo, el análisis de contenido de las entrevistas y sesiones posteriores identificaron las características del curso de capacitación que los evaluadores encontraron más valiosas, como los métodos de observación, las estrategias necesarias para superar los problemas de comunicación de los evaluadores con los profesores y la gestión de las sesiones posteriores. Los resultados también revelaron que el resultado del factor TE no se vio afectado por el curso de instrucción de capacitación.

Introduction

Evaluation is defined as the process of assigning merits or worth to the results of an observation or data collection (Cizek, 2000). In the area of English language teaching, Teacher Evaluation (TE) is considered a central concept since teaching and evaluation are intertwined in such a way that one cannot be discussed without touching the other. In the area of TE, there are many methods and instruments available, but classroom observation is often considered the main instrument, which is used extensively in many fields of education to evaluate a teacher's performance (Howard, 2010).

Regarding the high standing attributed to TE in teacher education, it is likewise essential to scrutinize those who evaluate teachers' performance, i.e., teacher evaluators. In order to gain proper results, teacher evaluators need special trainings in TE. As Stufflebeam (2001) puts it, proper evaluation training is essential to the health of the field. Teacher evaluation is not a haphazard entity and requires due training and experience. Schwandt (2008) argues that becoming a skilled evaluator is greater than the sum of the technical training elements. Evaluators' training is highly important because qualified teacher evaluators lead to qualified teachers and successful students. Additionally, the school principals and institute managers are in direct contact with society in general and with parents in particular. They must be responsive to the public, and this responsibility mostly lies on the teachers' shoulders. Generally, the procedure for TE at language institutes consists of evaluators observing classes and holding post-observation sessions with

---

teachers and the institute managers making decisions about the future status of teachers. The procedure is often summative and includes promotion, contract renewal or dismissal of teachers.

Therefore, the quality of the supervision and TE is a high stakes activity. It plays an important role in the success or failure of the teachers and thus, requires due attention. In pursuance of a method to maximize teacher evaluation effectiveness, one of the best ways is explicit instruction, a systematic and structured methodology to train teacher evaluators. Explicit instruction of evaluators is a key factor to ensure appropriate teacher performance and student achievement. It consists of a series of supports or scaffolds in favor of teacher evaluators where they are guided through a systematic procedure on how to evaluate teachers honestly and professionally.

To respond to this challenge, the aim of this study was to explore how teacher evaluators change as a result of a training course (explicit instruction) on TE. This was achieved by investigating the perceptions and practices of evaluators before and after the training course. To achieve more meticulous results, the evaluators explored via a triangulated research design consisting of TE questionnaires, interviews, and debriefing sessions. Teacher evaluation consisted of several dimensions, comprising its criteria, method, system, outcome, and evaluators themselves.

Thus, the research questions for this study were as follows:

1. Does explicit instruction about teacher evaluation affect the evaluators' perceptions, method, system, content, purpose, and outcome of evaluation?
2. How does the teacher evaluation training course affect the evaluators' practice?

Evaluation is an indispensable part of any educational program management and ideally should be designed to promote good classroom teaching and learning and thus, preparing evaluators well is crucial. Therefore, this study attempted to identify the impact of TE training course on the evaluators' perceptions and practices. The results of this study reveal important insights into the administration of evaluators' training courses and identifying the effects on teachers' performance and students' achievements.

Literature Review

The literature on teacher evaluation is divided into several categories including: models of teacher evaluation (Freeman, 1982; Gebhard, 1990; Knox, 2008; Wallace, 1991); self-evaluation and peer-evaluation (Mann & Walsh, 2013; Trotman, 2015); mentoring and coaching (Mercado & Mann, 2015); classroom observation (Howard, 2010); feedback from supervisors (Copland, 2010; Donaghue, 2015); explicit instruction of teachers (Hammond & Moore, 2018); students' evaluation of teachers (Greimel-Fuhrmann & Geyer, 2003; Taqi et al., 2014), and administrators' views on teacher evaluation (Maharaj, 2014). Compared to the number of studies on teacher evaluators and supervisors (Chen & Cheng, 2013; Peterson, 2004), fewer studies have been conducted to emphasize evaluator training (Dillman, 2012; Leahy, 2012; McClellan et al., 2012). More importantly, the perceptions and practices of evaluators as a result of a training course has rarely been explored and analyzed.

An exploration of literature on language teacher evaluation revealed that most of the studies have focused on classroom observation and the impact of observation on teachers' instructional behavior, meaning that what happens to the evaluators themselves has remained almost untouched. As long as our evaluators' preparation programs do not sufficiently prepare individuals to become qualified evaluators, a solution may be designing a program to train and improve evaluators because the few existing evaluator training studies do not delve into the perceptions and practices of evaluators simultaneously.

In a seminal study, Dillman (2012) explored the educational experiences of two groups of new and graduate student evaluators of English language teachers. The study examined what they believed had contributed to their evaluation competencies. The findings showed that coursework, fieldwork, and participation in professional activities like conferences, mentorship, and self-directed learning were perceived the most significant educational experiences.

In another study, Leahy (2012) reported on various models of teacher evaluator training programs utilized by states and local districts in the USA. In some states, master educators were hired specifically for the purpose of evaluating teachers. In others, online training was practiced, yet others presented three days training and testing of teachers who had served as observers for three years. Furthermore, McClellan et al. (2012) presented a practitioner series for teacher evaluator training and certification. The training format in this model was of two types: face to face and online. The components of observer training in this program consisted of bias training, observation process and tools, observation rubrics, and video trainings. The

program emphasized that the system should be transparent to the teachers observed so that they would have more confidence in the outcome. At the end of the program, there was a test and those who could pass the test received an observation training certificate. However, what Dillman (2012), Leahy (2012), and McClellan et al. (2012) failed to do was carry out explicit instruction about TE with evaluators and measure its effect on evaluators' perceptions and practices simultaneously, which was the purpose of this study.

One study that responded to this gap in the literature is that of Sweeney (1992) in which the effects of evaluator training on teacher education were determined. To this end, a 30-hour training course was carried out on 64 school principals and supervisors. In the training course, the participants learned how to collect valid data in the classroom, analyze performance, provide feedback, and coach for improved performance. The pretests and posttests completed by the trainers indicated that they had acquired the task-relevant knowledge of the evaluation concepts necessary to provide training for the evaluators. In other words, the trainers were taught instructional content to teach other supervisors in the future.

Regarding the evaluators' perceptions of teacher evaluation, King (2015) conducted a study in which the views of two appraisers toward TE were taken into account only through an interview. Based on the results, the appraisers identified evaluation useful for the experienced teachers, but at the same time stressful for other teachers. Regarding classroom observation, the appraisers deemed observation effective for reflective teachers, but not necessarily for other teachers. The usefulness of observation was questioned for quality assurance and management decisions because no change happened in some unsuccessful teachers. Sawchuk (2012) states that "it is not enough for observers to understand the theory and philosophy behind the observation instrument, they have to also be able to demonstrate accurate scoring across grade levels and subjects" (p. 2). Thus, theory is not sufficient, and training courses need to be practical.

The theoretical foundation used in this study for explicit instruction of the evaluators is Danielson's framework (2013 version). It consists of four domains: (1) planning and preparation; (2) the classroom environment; (3) instruction; and (4) professional responsibilities. The domains formed the basis for the training course employed in this study. This framework has been used to reflect a holistic picture of teaching practice, both inside the classroom (Domains 2 and 3) and outside the classroom (Domains 1 and 4). Domain 1 (planning and preparation) includes demonstrating knowledge of content and pedagogy, knowledge of students and resources, setting instructional outcomes, designing coherent instruction, and designing student assessments. Domain 2 (the classroom environment) includes creating an environment of respect and rapport and a culture for learning, managing classroom procedures and student behavior, and organizing physical space. Domain 3 (instruction) includes communicating with students, using questioning and discussion techniques, engaging students in learning, using assessment in instruction, and demonstrating flexibility and responsiveness. Domain 4 (professional responsibilities) includes reflecting on teaching, maintaining accurate records, communicating with families, participating in the professional community, growing and developing professionally, and showing professionalism. Each domain consists of four levels: unsatisfactory (level 1), basic (2), proficient (3), and distinguished (4).

As demonstrated above, there have been some studies on evaluator training; however, the literature lacks an empirical study to examine evaluators through a triangulated research design. Particularly, we lack evidence of how training courses affect the actual practice of evaluators in their classroom observations and in the post-conference debriefing sessions. Hence, this study contributes to the field by examining the perceptions and practices of EFL teacher evaluators and providing a detailed picture of the evaluators before and after a training program. In addition, it offers invaluable insights into the process of evaluators' classroom evaluation and training.

Methodology

In order to answer the research questions, this study used a mixed research design, consisting of qualitative and a quantitative design (Schoonenboom & Johnson, 2017). Accordingly, a teacher evaluation training course was designed, and its impact was evaluated through triangulated data collection techniques. The researchers were also the TE course designers and course instructors. To triangulate the findings, the present study used surveys through questionnaires, semi-structured interviews, recordings of debriefing sessions, and materials used in the training sessions.

Participants

The participants of the study consisted of twenty teacher evaluators working in private English language institutes in Tehran as part-time or full-time evaluators, and were thus, familiar with the basic criteria and

components of language teaching and teacher evaluation. The selection was based on their willingness and purposive sampling (Creswell, 2013). They were selected from among experienced teacher evaluators who had a Master's or Ph.D. degree in Teaching English as a Foreign Language (TEFL), and had orally announced their readiness to participate in the study, and seemingly, met the requirements of the research study. They were both males and females aged 22 to 32 with minimum five years of teacher evaluation experience. In the interview and post observation debriefing sessions of the study, all the participants agreed to participate.

<u>Instrumentation</u>

The instruments and materials used in this study consisted of a teacher evaluation questionnaire, semi-structured interviews, post-observation debriefing sessions, PowerPoint files, videos, worksheets, and handouts used in the training sessions.

Teacher Evaluation Questionnaire

The teacher evaluation questionnaire used in this study was designed by the researchers (Estaji & Shafaghi, 2018), and distributed among the 20 participants both before and after the training course. The TE questionnaire was used to examine the evaluators' perceptions toward evaluation of English language teachers' performance considering the various dimensions of teacher evaluation. It was comprised of 90 items on a 5-point Likert scale ranging from totally agree (rated 5) to totally disagree (rated 1). In order to measure the reliability of the questionnaire, some positive-worded items were negatively-worded so as to strengthen the reliability. Then, Cronbach's alpha was run. The results indicated a satisfactory level of reliability, i.e., $\alpha = .92$. Moreover, high correlations were found between the responses on each item and the whole questionnaire. Besides, the reliability of individual factors of the questionnaire was also estimated. The results showed all of the indices for the internal consistency of the factors were above .7, which indicates an acceptable level of internal consistency. Afterward, the validity of the questionnaire was ensured by exploratory and confirmatory factor analysis, demonstrating that the questionnaire consisted of a clear six-factor structure. The six constructs were the Perception, Method, System, Content, Purpose, and Outcome of teacher evaluation, which are briefly defined in the following:

1. **Method:** It refers to systematic procedures that evaluators take in their daily teacher evaluation practices.
2. **Outcome:** It shows the final product or consequences of evaluating teachers. Besides, it assists in making personnel decisions related to promotion or dismissal of the teachers.
3. **Perception:** It refers to the viewpoints and understandings of teacher evaluators toward teacher evaluation. It reflects how evaluators perceive and interpret the concept of TE.
4. **Purpose:** It reveals the purposes for evaluating teachers. The purpose for TE can range from overcoming teachers' instructional problems, to promoting teachers' knowledge of teaching methodologies, to seeking mutual compromise between teachers and evaluators, to improving students' achievements and other related factors.
5. **Content:** It reflects the essence of TE in general. It also involves the criteria that evaluators use to evaluate teachers such as teachers' skills in teaching and assessment, planning and classroom management, oral and written communication skills, and teacher's knowledge of the subject matter and curriculum.
6. **System:** It refers to the several organized steps in teacher evaluation. It reflects the process of assessing the performance of the entire TE system to discover how it is likely to perform in real workplace conditions.

Semi-Structured Interview

Two semi-structured interview sessions were carried out in the study, one before and one after the training course, each one taking about fifteen minutes. The interviews were used as complementary tools to get more robust data. The interview questions of both rounds were the same, with ten items developed by the researchers, which reflected upon the main constructs of teacher evaluation, such as the definition of teacher evaluation, their perceptions, instruments, methods, and strategies of teacher evaluation employed by the participants. To ensure content validity, three language experts were consulted concerning the interview questions, and their viewpoints on the main constructs of the questions were taken into account. This was to ensure the content and language appropriateness of the interview questions. In addition, a sample interview was conducted for piloting purposes.

Post-Observation Debriefing Sessions

The post-observation debriefing sessions are the sessions held after classroom observation between teacher and evaluator to elaborate on teacher performance and behavior, and help them improve their performance. In this study, the twenty participants' debriefing sessions with teachers were recorded once before and once after the training course one-on-one. The length of the debriefing sessions was not fixed, ranging from ten

to thirty minutes each. The aim was to explore the effect of the training course on the actual practice of the evaluators in the post-observation sessions.

Course Design & Materials

Based on the results derived from the initial analysis of the TE questionnaires, the semi-structured interviews, and the debriefing sessions, a teacher evaluation training course was conducted in an English language institute in Tehran. Only three out of the twenty participants of the study had already participated in a workshop on teacher evaluation. Therefore, it was decided to hold a four-session course to train the evaluators on various dimensions of teacher evaluation. The objective behind holding the training course was to introduce the concept of teacher evaluation from various perspectives, familiarize the participants with the current status of TE in the world, make the participants ready for the future teacher evaluations, introduce various models of TE, model and simulate the debriefing sessions between teachers and evaluators, and explore whether and to what extent explicit instruction would affect the performance of teacher evaluators.

The training course was in total twenty hours, held on four subsequent days each taking five hours with two short breaks. Getting insights from Danielson's (2013) TE framework, various materials such as PowerPoint files, handouts, worksheets, and video files on classroom observation, evaluation, and debriefing sessions were used. Similar to Danielson's framework, some video files were also used in the training course showing TE debriefing sessions. Additionally, TE tasks were designed in which the participants were asked to play the roles of teachers and evaluators to prepare for the post-observation debriefing sessions in real life situations. The idea of using tasks was derived from Danielson's framework emphasizing tasks that engage learners in deep learning. Furthermore, various types of TE models and forms were categorized for further discussion in the worksheets. The PowerPoint files also included topics such as the purpose and significance of the evaluation, duties of an evaluator, evaluation methods and systems, classroom observation, and the role of reflection in evaluation, all derived from Danielson's framework. The details of the four-session training course are described below.

In the first training session, the history of teacher evaluation and the major studies conducted on TE were described to the participants using PowerPoint files and detailed description. Likewise, there were discussions about the Perception, Method, System, Content, Purpose, and Outcome of teacher evaluation in the world. The participants shared their knowledge and experience with each other. Then mentoring and coaching were explicated with video clips.

In the second session, the practical aspects of teacher evaluation were explained to the participants. For instance, two types of evaluator bias were described: bias due to evaluator preferences and bias due to evaluator knowledge of the participants. Additionally, self-evaluation and peer-evaluation were introduced with examples. The participants were then divided into small groups, and started to practice self and peer-evaluation. Afterward, four different classroom observation videos were shown. Two videos depicted poor performance of two observers in the classroom observation sessions which resulted in dissatisfaction of both teachers and the observers. The other two video files showed qualified observers and successful classroom observations. Each video was replayed, and analyzed in detail, engaging all participants in the discussions.

The third training session began with introducing different sources of evidence for teacher evaluation. The sources included students' evaluation of teachers, administrators' views on teacher evaluation, evaluators' checklists, and the relationship between evaluators and teachers. Four sample video clips were then shown to the evaluators on debriefing sessions between teachers and evaluators, two in the form of group debriefing sessions and two individual debriefing sessions. The details of the video clips were discussed including the strengths and weaknesses of the evaluators and the tone and discourse of the evaluators. The task used was role play in which one evaluator took the role of a teacher, and the other took the role of an evaluator.

In the fourth session, the instruments that the evaluators used most for TE were examined. Most of the twenty participants had used outdated checklists as their evaluation tool in their classroom observations. The weaknesses of the checklists were discussed and a handout was given to the evaluators to describe the elements of an effective TE model. Afterward, the participants were asked to design their own TE checklists. Then the criteria for an effective TE were examined. "Since teachers become more conscious of the behaviors that the rubric considers desirable and effective, improved practice is often an attractive byproduct of this training" (McClellan et al., 2012, p. 7). At the end of the last session, the participants had the opportunity to freely express their opinions regarding various evaluation methods introduced in the training sessions.

Throughout the training course, Danielson's 2013 Teacher Evaluation Framework (TEF) was emphasized including its elements and levels.

Data Collection Procedure

After getting the participants' consent to cooperate in the study and piloting the research instruments and ensuring their reliability and validity, the TE questionnaire was distributed among the participants of the study. The questionnaires were administered to the participants, and clear explanations were provided to fill out the questionnaire. After the questionnaires were filled out, semi-structured interviews and debriefing sessions were held with all the participants of the study. Each participant was interviewed for fifteen minutes separately using a semi-structured interview protocol. The interviews and debriefing sessions were recorded on a Digital Voice Recorder (DVR). All the recorded interviews and debriefing sessions were transcribed, summarized, categorized, coded, and analyzed, and themes were extracted from the data. Afterward, the four-session training course on TE was held as described above.

After the training course, the three steps of data collection consisted of leading the participants to fill out the TE questionnaires, holding the interview sessions, and recording the debriefing sessions were carried out again for comparison purposes. The data collection procedure of the study is shown in Figure 1.

Figure 1: The data collection procedure of the study

Data Analysis

The first research question investigated if explicit instruction on evaluation affected the evaluators' Perceptions, Method, System, Content, Purpose, and Outcome of evaluation. In order to analyze the data pertaining to the first research question, after checking the normality assumptions using one-sample Kolmogorov-Smirnov tests and Shapiro-Wolf tests, a two-way repeated measures Analysis of Variance (ANOVA) was run. Likewise, to identify the difference and interaction of the factors, a post hoc analysis was run.

The second research question examined how the TE training course affected the evaluators' practice. To answer this research question, the researchers had to observe the evaluators' actual evaluation of teachers in the debriefing sessions after the training course and transcriptions were made. For data analysis, content analysis was used in four stages: condensation, code, category, and theme. First, the text was shortened while the core meaning was still preserved. The next step was to develop codes/labels for the condensed meaning units. Then, a category was formed by grouping together the codes that were related to each other. Finally, a theme was derived by interpreting the underlying meaning (latent content) of the text. An interview sample from one of the evaluators is presented here to show how the constructs were derived from the content analysis of the interviews:

The teacher evaluation systems working in private language centers are somehow similar to each other. They consist of classroom observations, checklists, and post-conference sessions in almost all the institutes.

Condensation: The TE systems are alike, consisting of classroom observations, checklists, and post-conference sessions.

Code: classroom observations, checklists, and post-conference sessions

Category: Teacher Evaluation Systems

Theme: In fact it's not a TE system; it's a norm which is uncritically practiced by almost all language centers.

Afterward, the post-observation debriefing sessions recorded before and after the training course were analyzed through content analysis of the concepts highlighted by the evaluators about teachers' performance. Additionally, the conversations between teachers and evaluators were analyzed through discourse analysis taking into account the tone, status, and language used by both parties.

Findings

Results for the First Research Question

The first research question examined if explicit instruction on evaluation would have any effect on evaluators in terms of Perception, Method, System, Content, Purpose, and Outcome of teacher evaluation. In this study, there were two independent variables each with different levels, namely, tests (pretest and posttest) and teacher evaluation factor (method, outcome, perception, purpose, content, system); therefore, one two-way repeated measures ANOVA was run so as to examine this research question. Figure 2 shows the participants' perception in the six factors of TE in the pretest and posttest.
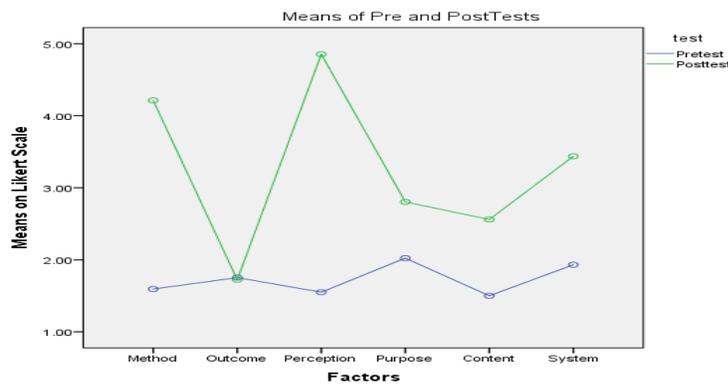

Figure 2: The participants' perceptions in the six factors in the pretest and posttest where 1.00 indicates 'totally disagree' and 5.00 indicates 'totally agree' on the Likert Scale.

The results of one-sample Kolmogorov-Smirnov test and Shapiro-Wolf test revealed that all data sets turned out to be normally distributed since the significance levels of all of them were greater than .05. Therefore, the two-way ANOVA could be run for them. Table 1 presents the descriptive statistics of the participants' perceptions across six factors in the pretest and posttest.

		N	Mean	Std. Error	Std. Deviation
Method	pretest	20	1.59	.01	.08
	posttest	20	4.21	.04	.18
Outcome	pretest	20	1.75	.04	.21
	posttest	20	1.72	.04	.21
Perception	pretest	20	1.55	.05	.24
	posttest	20	4.85	.01	.08
Purpose	pretest	20	2.02	.07	.33
	posttest	20	2.80	.08	.39
Content	pretest	20	1.50	.05	.22
	posttest	20	2.55	.07	.32
System	pretest	20	1.93	.09	.40
	posttest	20	3.43	.07	.32

Table 1: Descriptive statistics of the participants' perceptions in the six factors in pretest and posttest where 1.00 indicates "totally disagree" and 5.00 indicates "totally agree" on the Likert Scale.

As seen in Table 1, the participants outperformed in the posttest of Perception (note the 4.85 mean score in the posttest) compared to other factors, and did not perform much differently in the posttest of Outcome (1.75/1.72 mean scores). So as to see whether the aforesaid differences were statistically significant, a two-way repeated measures ANOVA was conducted. Before reporting the results of the ANOVA, the equality of variances and equality of covariances were checked by Levene's Test and Box' Test. The results revealed that no assumptions were violated. The results of the two-way repeated measures ANOVA are reported in Table 2.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared (effect size)
Test	142.26	1	142.26	1882.26	0.00	0.94
Teacher Evaluation	59.90	5	11.98	162.72	0.00	0.87
Test*Teacher Evaluation	75.44	5	15.08	199.64	0.00	0.89
Error(test)	8.61	114	0.07			

Table 2: The results of tests and components of teacher evaluation and the interaction between them

As seen in Table 2, the difference in the means of the pretest and posttest was statistically significant, i.e., $F(1, 114) = 1882.26$, $p < .05$. This difference was shown to be meaningful due to the high effect size ($\eta_p^2 = .9$). Cohen (1988) declares that the effect size below or equal to 0.01 is low, equal to 0.06 is moderate, and above or equal to 0.14 is large. Table 2 also indicates that the difference in the mean of the factors of teacher evaluation was significant, i.e., $F(5, 114) = 162.72$, $p < .05$, with a high effect size ($\eta_p^2 = .8$). The interaction between the factors and tests was also significant, i.e., $F(5, 114) = 199.64$, $p < .05$, with a high effect size ($\eta_p^2 = .8$). Since the difference in the means of the factors of teacher evaluation was significant, a post hoc analysis was run. Table 3 presents the results of the post hoc analysis.

(I) groups	(J) groups	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
Method	Outcome	1.16[*]	.06	.00	.98	1.34
	Perception	-.29[*]	.06	.00	-.48	-.11
	Purpose	.49[*]	.06	.00	.30	.67
	Content	.87[*]	.06	.00	.69	1.05
	System	.21[*]	.06	.00	.03	.40
Outcome	Perception	-1.46[*]	.06	.00	-1.64	-1.28
	Purpose	-.67[*]	.06	.00	-.85	-.49
	Content	-.29[*]	.06	.00	-.47	-.10
	System	-.94[*]	.06	.00	-1.12	-.76
Perception	Purpose	.78[*]	.06	.00	.60	.97
	Content	1.17[*]	.06	.00	.99	1.35
	System	.51[*]	.06	.00	.33	.69
Purpose	Content	.38[*]	.06	.00	.20	.56
	System	-.27[*]	.06	.00	-.45	-.08
Content	System	-.65[*]	.06	.00	-.83	-.47

Table 3: The results of the post hoc analysis for the factors of teacher evaluation

Table 3 indicates a statistically significant difference between each of the factors ($p > .05$). A significant interactive role of tests and factors was also obtained. Table 4 shows the difference in the means of the pretest and posttest for all factors.

Factors	(I) test	(J) test	Mean Difference (I-J)	Std. Error	Sig.[b]	Cohen's d	95% Confidence Interval for Difference[b]	
							Lower Bound	Upper Bound
Method	2	1	2.62[*]	.08	.00	91.52	2.44	2.79
Outcome	2	1	-.02	.08	.73	------	-.20	.14
Perception	2	1	3.30[*]	.08	.00	91.92	3.13	3.47
Purpose	2	1	.78[*]	.08	.00	10.37	.60	.95
Content	2	1	1.05[*]	.08	.00	17.26	.88	1.23
System	2	1	1.50[*]	.08	.00	17.48	1.33	1.67

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
[b] Adjustment for multiple comparisons: Bonferroni.

Table 4: The results of the post hoc analysis for the interaction of the factors and rests

Table 4 shows that the difference in the means of the pretest and posttest for all factors was significant except for Outcome. The most significant difference in the means of the pretest and posttest belongs to Perception (91.92), Method (91.52), System (17.48), Content (17.26), and then Purpose (10.37), respectively. Based on the results, the null hypothesis pertaining to the first research question was rejected. In other words, explicit instruction on evaluation does affect the evaluators in terms of Perception, Method, System, Content, and Purpose of teacher evaluation. Further research is needed to understand why Outcome was not affected by the training intervention.

Results for the Second Research Question

While in the first research question, the Perception, Method, System, Content, Purpose and Outcome of teacher evaluation were examined, in this research question the actual practice of teacher evaluators was examined after the training course. Based on the results, in the first debriefing sessions prior to the course, the evaluators mostly focused on the teachers' behavior in the classroom and the teaching skills and strategies they used; whereas, in the second debriefing sessions, changes as explained below were evident.

To begin with, the content of the debriefing sessions actually changed in the second debriefing sessions. In addition, new methods of observation and debriefing sessions were used by the evaluators after the training course. Moreover, the evaluators abandoned some habits of classroom observations such as unexpected visits to the classroom or interruption of teachers while teaching. The findings showed that the evaluators had put the theories into practice. For example, in the observation sessions, the evaluators focused on how and to what extent the students were engaged in learning. They started to observe the students' learning and teachers' performance simultaneously. The evaluators updated the classroom observation checklists by the guidelines they received in the training course and they added new items to describe all evidence in the classroom. The following excerpts represent the evaluators' remarks in interviews with the researchers, and the relevant themes extracted which arose after the training course in the second debriefing sessions.

A. Students' attitudes toward teacher

We did not use to take students' attitude into account for teacher evaluation purposes. However, after the training course, we've started to attend students' attitude toward teachers as one source of evidence. Students' learning has also been highlighted in our observation objectives. (Evaluator 15)

B. No interruption or unexpected visits

The majority of teachers I know feel frustrated by sudden observations and interruptions in their teaching. So, I have certainly changed my method of observation. (Evaluator 2)

C. Combining formative with summative evaluation of teacher performance

Summative evaluation focuses on the results of classroom observations and the amount of success achieved by the teacher. Formative evaluation, on the other hand, emphasizes the entire process from the beginning to the end. Thus, I think both should be incorporated into the evaluation programs. (Evaluator 9)

D. Methods of debriefing

The evaluation procedure I follow obeys the regulations of the institute. First, I do classroom observation; next, I let the teacher express his/her comments on the observation session, and then I provide the teacher with my suggestions and comments. (Evaluator 5)

Teachers should be evaluated in a standard and step-by-step procedure. Evaluation of teachers, like any type of evaluation, should occur based on a system and a procedure. After the training course, I followed the pre-

observation, classroom observation, and post-observation procedure emphasized in the course and also two observations were carried out per semester along with pre-and post-observations. (Evaluator 12)

E. Relation between teacher and evaluator

The atmosphere of debriefing sessions is often rigid. A solution for more effective evaluation is by building mutual relationships between teachers and evaluators. Informal chat between teachers and evaluators is an efficient way to break the ice between them which leads to high amount of mutual comprehension and friendship. (Evaluator 3)

After the training course, I had a sense that we and teachers intend to learn from each other rather than teaching each other. (Evaluator 8)

F. Teachers' sufficient time to express themselves in the session

Teachers should have the opportunity to start the debriefing session and make their own comments instead of taking a passive or defensive role. The amount of time in the debriefing sessions should be equally divided between the two parties. (Evaluator 15)

G. Peer evaluation and self-evaluation

Teacher evaluation is very helpful in one's personal and professional development. It makes teachers and teacher evaluators do self-reflection and peer-reflection activities by which they look at the profession from a new perspective. (Evaluator 2)

Teachers feel more comfortable when they do self-evaluation and peer-evaluation. Specially, novice teachers are energized when being observed by their peer teachers. (Evaluator 11)

H. Observing students' learning and teachers' evaluation simultaneously

To do classroom observation, I used to observe teachers in the classroom for years. But after the training, I pay attention to how students communicate with the teacher, how they are engaged in the process of learning, how they take turn, how they participate in the classroom activities and tasks, and how they are assessed in the class. I think evaluators must observe the students' learning and teachers' performance simultaneously. (Evaluator 7)

As for the debriefing sessions, the evaluators started with compliments about the strengths of the teachers and the class. Next, the evaluators asked questions about the points they did not understand, and finally, they made positively-worded suggestions for teacher improvements. The compliments and questions were presented explicitly whereas the suggestions were proposed implicitly using expressions such as "Don't you think" and "How about." The flow of conversations was smooth, and the ending was similar to the beginning, with a friendly and respectful mood. All in all, the results of this research question revealed the effectiveness of the training course in various aspects of teacher evaluation and how it challenged the evaluators' skills, perceptions, experiences, and practices of teacher evaluation.

Discussion

The present study investigated whether explicit instruction affected evaluators in terms of Perception, Method, System, Content, Purpose and Outcome of teacher evaluation. Additionally, it intended to explore how the training of the evaluators affected their practices in actual evaluation. The analysis of the results showed that the training course made a statistically significant difference in the evaluators in terms of the six mentioned factors. Among the six factors of teacher evaluation, determined in the TE questionnaire, the training sessions had the greatest effect on the Perception, followed by Method, System, Content, and then on Purpose; while it did not significantly influence Outcome. In particular, the perception of the evaluators was the most affected factor among all other factors. This result was also confirmed by the interview results.

The analysis of the results also gave rise to the new themes extracted in the second phase of data collection. For instance, in the follow-up interview sessions, the evaluators claimed to be involved in formative evaluation of teachers; however, before the training course, they were merely doing summative evaluation. Moreover, the analysis of the second post-observation debriefing sessions confirmed the evaluators' use of formative evaluation. This goes in line with Clenchy (2017), who argues that a strong system of teacher evaluation consisting of both formative and summative evaluation strategies is essential in order to enhance teacher performance and maximize student achievement.

Another important difference between the first and second observation was that in the second, the evaluators started to take the students' comments into account while evaluating teachers. Using students' comments and suggestions about the teacher and teaching was inserted into the evaluators' perceptions and practices. This is in line with Kane et al. (2011), who believe that engaging students in teacher evaluation helps to increase teacher effectiveness. However, the evaluation of teachers should not be merely

based on one source, the students, because, as Strauss and Corbin (1998) indicate, the students' perceptions might be based on their personal feelings toward the teacher rather than the teacher's performance.

The results also differed in that the evaluators, who merely observed teachers in the classroom for years, were now asked to observe the students as well. After the training course, the evaluators were sensitive to how the students communicated with the teacher, how they were engaged in the process of learning, how they took turns, how they participated in the classroom activities and tasks, and how they were assessed in the class. The perceptions of the evaluators also changed because they considered teacher evaluation in terms of professional development (PD). In the second administration of the TE questionnaires and the second series of interviews, the evaluators discussed doing evaluation not only for the sake of evaluating teachers but also for their own PD. Mann and Walsh (2013) prefer the term continuing professional development (CPD). Roberts (1998) also puts forward a strong argument that PD is only possible through a reflection process that puts self-monitoring and self-evaluation at the heart of things; he views these processes as "the only possible basis for long-term change" (p. 305).

Howard (2015) recommends teachers assume more participatory roles in evaluation in order to give them greater voice, and promote their professional development. Likewise, Wang and Day (2001) claim that if supervisors give a participatory role to the teachers in the process of supervision and evaluation, they are more likely to change the focus of classroom observations from a means of teacher evaluation to teacher development. In terms of method, the training course made the participants take advantage of self-evaluation. An important method of self-evaluation is through video recording which enables the teachers to think of alternatives to teaching practices and professional development, and forms the beliefs about the relationship between teaching behavior and student learning (Gebhard & Oprandy, 1999). The evaluators' perceptions changed as they declared one observation during each semester was not sufficient; rather, two to three observation sessions were required to render a more accurate evaluation of teachers. This complies with Myung and Martinez (2013), who argue that a reliable assessment of teaching requires multiple observations.

Regarding the second research question, which explored the effect of the training course on the evaluators' actual practice, the analysis of the results stemming from the comparison of the content of the first debriefing sessions with the second ones verified the effectiveness of the training course in the teacher evaluators' performance. The results showed that the evaluators actually practiced what they had theorized after the training course. Since the training course was held based on the Danielson's (2013) TE framework, the evaluators were trained in terms of: (a) planning and preparation (designing new checklists for TE, avoiding interruption and unexpected classroom visits, holing pre observation, observation, and post observation sessions sequentially); (b) the classroom environment (observing students as well as teachers, examining students' learning and their relation with teacher); (c) instruction (content, methods, and strategies of teaching); and (d) professional responsibilities (reflecting on teaching, keeping accurate records of classroom observations and debriefing sessions, and doing peer-evaluation and self-evaluation).

Moreover, the evaluators had their classroom observation checklists updated to include new themes. The analysis of the results also revealed that the evaluators' perceptions shifted away from criticizing teachers in the post-observation debriefing sessions to the analysis of the entire events involved in the classroom observation sessions. The TE checklists changed because almost all the evaluators used to utilize outdated checklists for classroom observations. In their checklists, the evaluators were free to add open-ended items. The methods the evaluators applied in the classroom observation sessions also changed after the training course. The evaluators had learned to sit in a position to be able to see the teachers along with all the students. In other words, the evaluators had learned to observe the entire class and not just the teacher. The evaluators had become conscious of observing the relationship between the students and teacher. They started to apply the three-phase procedure of pre-observation, observation, and post-observation model in the quality control (QC) section of their workplace. There were no longer unexpected visits to the classroom by the evaluators because the objective of teacher evaluation had changed in the quality control center of the language centers. This is in line with domain 1 in Danielson's (2013) TE framework. Besides, it complies with Howard (2015), who suggests that teachers who have a pre-observation conference session with their observers do feel that they have a voice in the evaluation process.

One another change was that the number of debriefing sessions grew because the evaluators realized that only one debriefing session would not suffice. The norm had changed to two observations per semester along with one pre-observation debriefing session. The content of the debriefing sessions had also changed.

The issues covered in the debriefing sessions contained new important themes: for example, the teachers' cooperative approach toward students' parents and other colleagues. Overcoming the students' behavioral problems was another issue emphasized. In the debriefing sessions, the teachers had the opportunity to take the initiative in running the session, and make their own comments instead of taking a passive or defensive role.

Furthermore, the evaluators made modifications in their relationship with the teachers. There was no longer any trace of interruption by the evaluators in the middle of the teachers' talk. The evaluators allowed the teachers to explain their teaching process and present their ideas about the class and students. The evaluators and teachers held formal conversations with each other with a supportive tone. The evaluators assigned the main part of the debriefing sessions to highlighting the strengths of the teachers, and encouraging them in their career and just after the admiration, they talked about the weak points with a mild and respectful tone. In addition, the teachers started to employ various strategies such as holding meetings with students, parents, and the personnel; they also explained the objectives of the course clearly.

The evaluators had learned not to look down on the teachers, and instead, strengthen their relationship with them. In the debriefing sessions, the evaluators started to have informal talks with the teachers in addition to the formal ones. They played the role of a facilitator in reflecting the teachers' problems, and tried to give promotions to the qualified teachers. Another important aspect added to the second debriefing sessions was the students themselves and their learning. In the debriefing sessions, every note taken by the evaluators in the observation session was examined as evidence. They included the students' attitude toward the teacher and the class, the students' pace of learning English, the way they were involved in the classroom discussions, assessment of the students, parents' feedback, and the time and effort students took in the class.

Conclusion and Implications

Evaluators are the cornerstones of evaluation, and need professional training to be successful in their jobs. LaVelle and Donaldson (2010) argue that "evaluators are made, not born, and an extended period of training is necessary to master the evaluation-specific skills and knowledge necessary to provide quality service to clients, and be socialized into the professional frameworks, standards, and ethical guidelines" (p. 10). According to Bryk and Schneider (2002), identifying the teachers' weaknesses is a central step in supporting their improvement. However, some teachers might not be interested in getting feedback from the evaluators. Therefore, finding the right method to relate with teachers is vital for the evaluators, and this study yielded satisfactory results in this regard.

The results of this study indicated that teacher evaluation which is currently practiced in Iran's language centers suffers from various shortcomings. Shifting the focus of teacher evaluation from summative evaluation to a multi-dimensional formative evaluation, considering the whole picture of teachers' strengths and weaknesses, and focusing on "how is a teacher effective" would help evaluators to have a more accurate picture of teacher effectiveness (Kraft & Gilmour, 2017). Formal training of teacher evaluators is crucial since evaluation and judgement of teacher performance should be reliable, valid, and fair, and in line with the professional development of teachers.

The process of teacher evaluation described in this study would offer both teachers and their evaluators the opportunity to work together to improve teacher performance. In order for a teacher evaluation system to be effective, an overall reexamination of teacher evaluation methods would be required. Regular debriefing sessions between teachers and teacher evaluators would assist teachers in regulating their pedagogical practices. Teachers who are engaged in reflective practices and self-assessment can make appropriate decisions, and better monitor the students' performance (McMeniman et al., 2003).

Nevertheless, just like other studies, the current study is subject to some limitations. The reason can be traced in the limited size of the sample and the limited time spent on each interview and the training course. Similar qualitative and quantitative studies in the teacher evaluation field would supplement this research, and offer greater insight into how teacher evaluation could develop and contribute to teacher growth. If teacher evaluation is well planned and executed according to the standards of best practice, the results will be improvement in teaching, students' performance, and satisfaction of all stakeholders (McClellan et al., 2012). Other research studies might explore the way evaluators perceive the impact of a web-based training program on their evaluation planning and practices. They can also take advantage of teachers' perceptions to update the criteria for teacher evaluation. Besides, teachers' self-assessment can also be considered as a good topic for conducting further studies on TE. Similar studies can utilize task-based evaluator training

to examine if it can promote the evaluation quality. Finally, further research can also include interviews with the teachers who have undergone the evaluations, and investigate whether they have noticed the changes that have been reported in the present study.

References

Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* Russell Sage Foundation.

Chen, C. W., & Cheng, Y. (2013). The supervisory process of EFL teachers: A case study. *The Electronic Journal for English as a Second Language (TESL-EJ, 17*(1), 51-72. https://www.tesl-ej.org/wordpress/issues/volume17/ej65/ej65a1

Cizek, G. 2000: Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice, 19*(2), 19-23. https://doi.org/10.1111/j.1745-3992.2000.tb00026.x

Clenchy, K. R. (2017). *Teacher evaluation models: Compliance or growth oriented?* [Unpublished doctoral dissertation]. Northeastern University. https://repository.library.northeastern.edu/files/neu:cj82gm547/fulltext.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Copland, F. (2010). Causes of tension in post-observation feedback in pre-service teacher training: An alternative view. *Teaching and Teacher Education, 26*(3), 466-472. https://doi.org/10.1016/j.tate.2009.06.001

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches.* Sage.

Danielson, C. (2013). *The framework for teaching evaluation instrument.* Danielson Group.

Dillman, L. M. (2012). Evaluator skill acquisition: Linking educational experiences to competencies. *American Journal of Evaluation, 34*(2), 270-285. https://doi.org/10.1177%2F1098214012464512

Donaghue, H. (2015). Differences between supervisors' espoused feedback styles and their discourse in post-observation meetings. In A. Howard & H. Donaghue (Eds.), *Teacher Evaluation in Second Language Education* (pp. 117-134). Bloomsbury Academic.

Estaji, M., & Shafaghi, M. (2018). Teacher evaluation in EFL context: Development and validation of a teacher evaluation questionnaire. *Issues in Language Teaching, 7*(2), 147-187*.* http://dx.doi.org/10.22054/ilt.2020.47348.433

Freeman, D. (1982). Observing teachers: Three approaches to in-service training and development. *TESOL Quarterly, 16*(3), 21-28. https://doi.org/10.2307/3586560

Gebhard, J. G. (1990). *The supervision of second and foreign language teachers* (EDO-FL-90-06)*.* ERIC Clearinghouse on Language and Linguistics. https://files.eric.ed.gov/fulltext/ED324971.pdf

Gebhard, J. G., & Oprandy, R. (1999). *Language teaching awareness: A guide to exploring beliefs and practices.* Cambridge University Press.

Greimel-Fuhrmann, B., & Geyer, A. (2003). Students' evaluation of teachers and instructional quality-analysis of relevant factors based on empirical evaluation research. *Assessment and Evaluation in Higher Education, 28*(3), 229-238. https://doi.org/10.1080/0260293032000059595

Hammond, L., & Moore, W. M. (2018). Teachers taking up explicit instruction: The impact of a professional development and directive instructional coaching model. *Australian Journal of Teacher Education, 43*(7), 110-133. https://doi.org/10.14221/ajte.2018v43n7.7

Howard, A. J. (2010). *Teacher appraisal: The impact of observation on teachers' classroom behavior* [Unpublished doctoral dissertation]. University of Warwick. http://wrap.warwick.ac.uk/3728

Howard, A. (2015). Giving voice to participants in second language education evaluation. In A. Howard & H. Donaghue (Eds.), *Teacher evaluation in second language education* (pp. 193-210). Bloomsbury.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587-613. https://doi.org/10.3368/jhr.46.3.587

King, M. (2015). Evaluating experienced teachers. In A. Howard & H. Donaghue (Eds.), *Teacher evaluation in second language education* (pp. 167-179). Bloomsbury.

Knox, L. (2008). *Three-step sequences in trainee teacher-supervisor talk: Mitigation and ambiguity in post-observation conferences* [Unpublished master's thesis]. University of Edinburgh.

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the Widget Effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher, 46*(5), 234-249.

LaVelle, J. M., & Donaldson, S. I. (2010). University-based evaluation training programs in the United States 1980-2008: An empirical examination. *American Journal of Evaluation, 31*(1), 9-23. https://doi.org/10.1177/1098214009356022

Leahy, K. (2012). *Teacher evaluator training: Ensuring quality classroom observers.* Education Commission of the States. www.ecs.org/clearinghouse/01/01/14/10114.pdf

Maharaj, S. (2014). Administrators' views on teacher evaluation: Examining Ontario's teacher performance appraisal. *Canadian Journal of Educational Administration and Policy, 152*(2), 1-58. https://journalhosting.ucalgary.ca/index.php/cjeap/article/view/42859/30716

Mann, S., & Walsh, S. (2013). RP or RIP: A critical perspective on reflective practice. *Applied Linguistics Review Journal, 4*(2), 291-315. https://doi.org/10.1515/applirev-2013-0013

McClellan, C., Atkinson, M., & Danielson, C. (2012). *Teacher evaluator training and certification: Lessons learned from the Measures of Effective Teaching project*. Teachscape. http://www.teachscape.com/resources/teacher-effectiveness-research/2012/02/teacher-evaluator-training-and-certification.html

McMeniman, M., Cumming, J., Wilson, J., Stevenson, J., & Sim, C. (2003). *Teacher knowledge in action: The impact of educational research*. Department of Education, Training and Youth Affairs.

Mercado, L. A., & Mann, S. (2015). Mentoring for teacher evaluation and development. In A. Howard & H. Donaghue (Eds.)*Teacher evaluation in second language education* (pp. 35-54). Bloomsbury.

Myung, J., & Martinez, K. (2013). *Strategies for enhancing the impact of post-observation feedback for teachers.* Carnegie Foundation for the Advancement of Teaching. https://www.carnegiefoundation.org/wp-content/uploads/2013/07/BRIEF_Feedback-for-Teachers.pdf

Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin, 88*(639), 60-79. https://doi.org/10.1177%2F019263650408863906

Roberts, J. (1998). Action research for language teachers*. Language Teacher Education, 13*(2), 92-106.

Sawchuk, S. (2012, March 14). Training of teacher-evaluators examined. *Education Week: Teacher Beat*. http://blogs.edweek.org/edweek/teacherbeat/2012/03/paper_examines_training_of_eva.html

Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 69*(2), 107–131.

Schwandt, T. A. (2008). Educating for intelligent belief in evaluation. *American Journal of Evaluation, 29*(3), 139-150. https://doi.org/10.1177/1098214008316889

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage.

Stufflebeam, D. L. (2001). Interdisciplinary Ph.D. programming in evaluation. *American Journal of Evaluation, 22*(2), 445-455. https://doi.org/10.1016/s1098-2140(01)00155-2

Sweeney, J. (1992). The effects of evaluator training on teacher evaluation. *Journal of Personnel Evaluation in Education, 6*(3), 7-14. https://doi.org/10.1007/bf00126915

Taqi, H. A., Al-Nouh, N. A., Dashti, A. A., & Shuqair, K. M. (2014). The perspectives of students and teachers in the English department in the college of basic education on the student evaluation of teachers. *Journal of Education and Learning, 3*(4), 71-89. https://doi.org/10.5539/jel.v3n4p71

Trotman, W. (2015). Reflective peer observation accounts: What do they reveal? In A. Howard & H. Donaghue (Eds.), *Teacher evaluation in second language education.* Bloomsbury Academic.

Wallace, M. J. (1991). *Training foreign language teachers: A reflective approach.* Cambridge University Press.

Wang, W., & Day, C. (2001, February). *Issues and concerns about classroom observation: Teachers' perspectives* [Conference session, FL027422]*.* Annual Meeting of the Teachers of English to Speakers of Other Languages (TESOL), St. Louis, MO, USA.http://files.eric.ed.gov/fulltext/ED467734.pdf