Investigating Iranian EFL Instructor Evaluation Scheme from End-Users' Perspective: Self-Evaluation vs. Students' Ratings

Fateme Naderian
English Department, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran
Email: art.naderian@gmail.com

Azizeh Chalak*
English Department, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran
Email: achalak@khuisf.ac.ir

Ahmad Ali Foroughi Abari
Educational Sciences Department, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran

Hossein Heidari Tabrizi
English Department, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran
Email: heidaritabrizi@gmail.com

Abstract

This study investigated Iranian EFL instructor evaluation scheme from end-users' perspective: self-evaluation vs. students' ratings. To do so, in the second semester of 2015-2016, 60 instructors and 1000 students of English Department of Islamic Azad University Isfahan (Khorasgan) Branch (IAUIB) were selected as those from whom the corpus of the study was extracted. The corpus was provided by administrating two rating scales online via the university website on each person's profile. Then, the results of their completed evaluation rating scales were compared. The study was accomplished through a non-experimental descriptive correlational design. The results revealed that almost no relationship was found between Iranian EFL instructors' self-evaluation and those done by their students at IAUIB. This study could benefit Iranian educationalists, policy makers, and evaluators in making informed pedagogical decisions and conducting more efficient teacher evaluation in English education of Iran.

Keywords: Iranian EFL Context, Rating Scale, Self-Evaluation, Students' Ratings, Teacher Evaluation Scheme

1. Introduction

Teachers as the heart of each educational system are supposed to help students to reach their full potentials and be prepared to lead a successful and productive life. Those students are successful who are properly taught. Therefore, the presence of effective teachers is a pressing need for a truly excellent educational system and when talking about higher education, the importance of this issue will be more. It is supported by some research such as a large-scaled empirical study conducted by Clotfelter, Ladd, and Vigdor (2010) that students' achievements are associated with the teachers' effectiveness. Therefore, it is considered a critical element in improving the students' achievements and one of the strong factors that plays an important role in achieving high-quality learning outcomes.

English instructors teaching in different branches of English major such as Translation Studies and Teaching English as a Foreign Language are not exceptions. In learning the English language successfully, lots of factors rely on EFL teachers and their role is very crucial because EFL students have few opportunities to apply what they learn in their daily life and in the real world. Consequently, the educational systems are persuaded to conduct the teacher effectiveness evaluation annually in order to identify and employ qualified EFL teachers, help them grow, and finally control whether their goals are met or not.

On the other hand, by integration of technology into the educational system, especially in the recent years, big changes have been observed in the field of language learning. These changes lead to some beneficial points such as facilitating accomplishment of pedagogical goals and providing more fulfilled atmosphere for students. Therefore, some higher education institutions, mainly the universities which are sure about the accessibility of their instructors and students try to implement them for evaluating their EFL instructors. Replacing these methods with traditional ways of the evaluation helps to solve the problems of time, cost, and place constraints in collecting the data for this purpose.

The teacher evaluation as a method for measuring the instructors' effectiveness is heavily relied upon. However, the reality of its findings remains undetermined. Moreover, with increasing discussion of using different rating scales as the measure to determine whether or not an instructor is teaching well, there is a big burden on the university

authorities to conduct the teacher evaluation effectively. Therefore, solving this problem and identifying the form of evaluation which can be helpful in achieving this goal needs more investigations. As a result, the evaluators may conduct less subjective evaluations with less human deficiencies, strengthen those instructors who are weak and further develop the skills of those who are already proficient.

2. Literature Review

EFL teaching is a multifaceted activity with several dimensions. Actually, it is considered a complex endeavor because an English classroom is a milieu in which EFL teachers have both educational and social responsibilities. In some studies such as (Bascia & Jacka, 2001; Roessinngh, 2006), EFL teachers have been referred to as life lines for their students because of the comprehensive nature of supports which they provide for them. Therefore, implementing an efficient monitoring and evaluation can entail in a continuous improvement of the effective teaching in the higher education.

Many potential roles have been suggested for English language teachers by several researchers such as (Harmer, 2001; Tudor, 1993). For example, Richards and Rodgers (2000) assume the role of English language teachers in association with the degree of the teachers' responsibility for determining the content of the course, their control over the amount of successful learning by students, and types of functions that the teachers are expected to fulfill. In addition, Wren (2006) presented a list of eight items on the attributes of a good teacher. He believes that an expert teacher is purposeful, uses instructional strategies, do not waste time, keeps the students actively engaged, creates a rich learning environment, uses data to inform the instruction, is in constant connection with students and their families, and has positive encouraging personality. All of these roles should be considered in the process of evaluation.

Angelo (1996) concentrated on the importance of teaching context in the process of teacher evaluation. Stake (1987) explained that in an effective evaluation of teaching, institutional goals, classroom environments, students' achievements, administrative organization and operations, curricular content, and impact of the program in the society should be studied concurrently because it is only in the context that teaching can be properly judged otherwise, the efforts would be invalid.

Some other researchers (such as Boggs, 1999; Ellett, Loup, Culross, McMullen, & Rugutt, 1997) also pointed out the common concern that there is a necessity to focus on the student learning in the evaluation of teachers more. With regard to the purposes of teacher evaluation, Danielson, and McGreal (2000) mentioned six different items. They emphasized that an effective evaluation process should be able to screen out unsuitable candidates, cut off cooperation of incompetent teachers, provide constructive feedback, distinguish and fortify outstanding efforts, provide staff with developmental directions, and unify the whole body of educational personnel around improved student learning.

According to Scriven, Micael, Barbra, and Susan (1987), there are two types of evaluation depending on the purpose which it is set for. If it is used for reshaping ongoing teaching activities, it is called formative evaluation and can be applied for measuring any variable in an English class. On the other hand, if the goal of evaluation is to classify evaluated variable into efficient and non-efficient it is a summative one. Isore (2009) states that "conducting a summative assessment is the most visible and recognizable way to evaluate someone, which consists of providing summary statements of teachers' capabilities through examinations, in order to measure aptitude and knowledge…" (p. 6). Furthermore, he defines formative teacher evaluation as "a qualitative appraisal on the teacher current practice aimed at identifying the strengths and weaknesses and providing adequate professional development opportunities for the areas in need of improvement" (p. 7).

Different systems depending on their policies, equipment, and time apply different methods or indicators to accomplish a comprehensive and fair evaluation. The information can come from student ratings which have been one of the most popular tools for measuring the teaching effectiveness in the higher education settings. They provide a collection of scores or values achieved from the questionnaires that are in turn completed by the students during each academic semester. According to Hinchy (2010), the content of the rating scales should be specifically concentrated on the items related to the teachers' practices, holistic aspects of the instruction, and interactions between the teachers and students which are key elements of educational environments. As a result, a rich and valuable source of information would be extracted. Theall, Abrami, and Mets (2001) defined five key functions for students' ratings. They mentioned that this instrument serves as a tool for the instructional improvement, as an evidence for promotion and tenure

decisions, as means which helps students in course selection, as one criterion for measuring program effectiveness, and as the continuing focus of active research and intensive debate.

In the later method, all of the information comes from teachers themselves. They complete a self-evaluation questionnaire and in this way express their own views about their teaching. According to Peterson (2000) and Kennedy (2005), involving teachers themselves in the process of evaluation is an essential because of two reasons first to gain their agreement on the process and then to enhance their performance. Sometimes, portfolios including teaching materials, samples of students' works, lesson plans, and a collection of what the teacher is doing in the classroom are used to complement self-evaluation. Interview with teachers in order to ask them about their skills and needs in highly structured format or not, questionnaires and surveys that are completed by those who are in continuous interaction with the teacher such as supervisors, principals, or students to testify teaching quality, peer reviews in which the instructors' knowledge and instruction are evaluated by other instructors or colleagues who share common knowledge of a specific discipline, students' interviews, and teachers' tests used for assessing their pedagogical skills are some other instruments that may be applied for collecting the data in the process of teaching evaluation.

Teacher evaluation studies have been under the spotlight of many researchers. In one of these studies conducted by Harvey and Barker (1970), the relationship between students' subjective judgments about their instructors' teaching effectiveness as a kind of personal conclusion, and their responses to a rating scale was investigated. In order to collect the data, the students were asked to evaluate two of their instructors who are considered as the most and the least effective one on the same basis using a questionnaire and a rating instrument. After that, items of the rating scale were correlated with the criterion of classification as most or least effective teacher and the criterion of summated ratings using a bilateral coefficient of correlation. The results of the study showed that the correlations were significantly different from zero correlation.

In another study, Brandt, Mathers, Oliva, Brown-Sims, and Hess (2007) studied policies used in Midwest region of the US through surveying 218 school districts with 140 participants. Their focus was on the way that the teacher evaluation results were reported

and used over there. They found that in Midwest, administrators use evaluations for the summative reporting and not for professional growth, in other words, it is used to decide whether to release or retain new teachers.

A relatively different study was carried out by Jacob and Lefgren (2008) in which 201 teachers were evaluated by principals on dedication, work ethic, classroom management, positive relationship with administrators, and ability to raise students' achievements. In addition, some data about themselves such as age, experience, educational attainment, and certification information as well as the students' achievement data were examined. Their purpose was to determine if these teachers, effective at increasing students' achievements, can be identified by administrators or not. The results of the study revealed that evaluating teachers by principals is an effective method for identifying the best and worst teachers, but they failed to distinguish teachers in the middle of the distribution.

Webb and Norton (2003) demonstrated that the level of effectiveness of an evaluation in providing influential feedback is more when it is accomplished by students rather than by teachers. In a similar study conducted by Centra (2003), it was revealed that there is little correlation between students' grades in a course and the results of students' ratings. Therefore, he clearly concluded that giving higher grades and less course work is not likely to improve the teachers' evaluation results of their student ratings. On the other hand, according to Meilke and Frontier (2012), perceptions and attitudes of teachers who participate in the process of the teacher evaluation is a key to success in increasing the potentiality of conducting an evaluation which leads to improved instruction and professional growth.

In a study by Ahmadi and Sajadi (2009), teachers were evaluated in order to decide whether the language department (LD) teachers are better to teach English for medical purpose or the discipline-specialist (DS) teachers. They used three questionnaires as the instrument for the data collection. The questionnaires were answered during the academic year of 2006-2007 by some vice-deans, some heads of the discipline-specialist departments and language departments as well as some students of English classes in six medical universities. After analyzing the data, they found that heads of language departments and students preferred LD teachers while the heads of discipline-specialist departments preferred DS teachers for teaching English for the medical purpose. Therefore, they

concluded that LD teachers should increase their knowledge of discipline while DS teachers should enhance their knowledge of language teaching.

In another study conducted by Rigi, Ghaderi, and Salimi (2016) teachers' performance assessment via audio-video data was compared with other methods used for teacher evaluation. The required data was collected by a researcher-made questionnaire administered to 142 teachers teaching in first grade high school and semi-structured interviews with eight science teachers who were selected with a random clustery method. Their findings showed that question centered approach is more appropriate for the current evaluation system and the next priorities are taken by accountability and social advocacy approaches. Moreover, they found that utilizing audio-video data in evaluating teachers' performance is helpful for evaluators in achieving their goals and solving the problems of the current system.

In other countries, the teacher evaluation has been an interesting and researched topic for many years that draws researchers' attention to be worked on and applied in educational settings (Barton & Shana, 2010; Jacob & Lefgren, 2008; Palmer & Marra 2004) in the hope of increasing students' success after graduation. However, less work has been conducted in Iran pertaining to this issue. Therefore, due to the scarcity of research studies conducted in Iran on evaluating teachers, especially in the EFL context, the following research question was raised:

Is there any relationship between the results of Iranian EFL instructors' self-evaluation and those done by Iranian English students at IAUIB?
To this end, the study was an attempt to investigate consistency between the results of two different rating scales used in evaluating EFL instructors.

3. Methodology

3.1. Research Design and Setting

The present study employed a non-experimental descriptive correlational design to explore those pieces of data gathered by the rubrics or rating scales in order to analyze the relationship between two already occurred variables. The study was conducted in the second semester of the academic year of 2015-2016 at the faculty of foreign language of IAUIB which is a branch of IAU located in Iran, Isfahan. Moving up in the way of being recognized as one of the supreme universities throughout the country was the main

justification for selecting IAUIB as the context of the study on the EFL instructor evaluation scheme. Another reason for choosing this university was easy accessibility to the corpus for the researcher and her personal contacts with those who could help with the data collection.

3.2. Corpus of the Study
Students' Ratings and Teachers' Self-Ratings

The students' rating scale contained three domains in which different aspects of the instructors' instructional activities were taken into consideration. The first domain labeled as teaching methodology consisted of four items with five-point Likert scale that elicited students' opinions about issues related to the instructors' teaching ability, methodology, and overall perceptions about the course, whereas its second section contained six items that exclusively elicited the information about instructor-student relationship, rapport, or ethics. The next domain was allotted to the ways adopted for assessing and evaluating the students. This rating scale was administered by the university to almost 1000 male and female students studying different English majors at different degrees during the academic year of 2015-2016. Table 1 shows their numbers.

Table 1

The Number of EFL Students at Different Majors in English Department of IAUIB

Department	Major	Degree	Number
	English Translation	B.A.	599
	TEFL	M.A.	101
English Language Department	Translation Studies	M.A.	97
	General Linguistics	M.A./Ph.D.	84
	TEFL	Ph.D.	16
Total			897

The teachers' self-rating scale was developed exactly the same as the students' rating scale, but there was one key difference. It was containing one extra domain with three items on the instructors' executive co-operational issues. Table 2 summarizes the EFL instructors' demographic characteristics; those who completed the rubric at the time of conducting the study.

Table 2

Demographic Backgrounds of EFL Instructors (n=60)

Demographic Factors	Labels	N
1. Gender	Male	35
	Female	25
2. Age	26-33years	13
	34-41years	22
	42-49 years	10
	50-57 years	7
	58-65 years	2
3. Educational Degree	M.A.	40
	Ph.D.	20
4. Academic Rank	Associate Professor	3
	Assistant Professor	12
	Instructor	9
	With No Ranks	36
5. Type of Cooperation	Full-Time	16
	Part-Time (Invited)	44

As Table 2 indicates, 16 full-time faculty members and 44 part-time instructors with different academic ranks ranging from associate professors, assistant professors, and instructors to those who worked as invited teachers with no ranking made, holding M.A. or Ph.D. degrees were teaching in different English majors at the English Department of IAUIB and were the target audience of the study. All were native speakers of Persian, although some of them experienced living in an English speaking or European country. With regard to the gender and age, the faculty members were 35 male and 25 female instructors whose age ranged from 26 to78.

3.3. Data Collection Procedure

All of the faculty members and students teaching and studying in different undergraduate and graduate programs of the English Department were selected through an accessible non-probabilistic cluster sampling method as those from whom the corpus of the study was extracted in order to investigate the Iranian EFL instructor evaluation scheme. The data collection was conducted by the authorities of the university online, via the university website, and through administrating the rubrics or rating scales on each instructor and student profile at the end of the second semester in the academic year of 2015-2016. The tabulation of the obtained data in a numerical form was also accomplished by the expert authorities of the university. In this way, a quick data gathering was provided. Then, the raw data collected from the students' ratings and teachers' self-ratings were ready to be analyzed in order to see how they are related to each other.

3.4. Data Analysis Procedure

After completion of the collection of the required data, in exploring the relationship between the data obtained from EFL students' ratings and their instructors' self-ratings, the raw data were fed into computer and the statistical calculations were run by the Statistical Package for Social Sciences (SPSS software) version 24 and the Pearson rank order correlation formula was identified as the best statistical method to analyze the data to see whether there was any significant difference between them or not. To this aim, the score of each instructor in each domain of teachers' self-ratings in addition to the mean score of the whole domains were correlated with the parallel scores on the students' ratings.

4. Results

This section presents the results of the data analysis stage of the study to help find answers to the research question. As it was mentioned above, it was aimed to unfold whether there was any relationship between the results of the Iranian EFL instructors' self-evaluation and those done by the students at the IAUIB. In an attempt to find an answer to this research question, Pearson correlation was conducted between Iranian EFL instructors' self-evaluation (and its subcomponents) and their students' evaluations of instructors (and its subcomponents). The results obtained from this test are shown in Tables 3 and 4:

Table 3

Results of Descriptive Statistics for Teachers' Self-Evaluations and Their Students' Evaluations

		N	Minimum	Maximum	Mean	*Std.* Deviation
Teachers' Evaluation	Evaluation Procedures	77	12.00	20.00	19.19	1.07
	Teaching Method	77	12.00	20.00	19.24	1.00
	Behavior	77	12.00	20.00	19.14	1.13
	Cooperation	77	12.00	20.00	18.62	1.39
	Teachers' Mean	77	12.00	20.00	19.01	1.00
Students' Evaluation	Evaluation Procedure	77	.00	19.58	17.82	2.12
	Teaching Method	77	.00	19.57	17.55	2.14
	Behavior	77	.00	19.67	17.93	2.14
	Students' Mean	77	.00	19.61	17.78	2.12

Prior to examining the Pearson correlation, it is worth taking a look at the descriptive statistics (Table 4) to find that the instructors' and students' mean scores differed for evaluation procedures, teaching methods, and behavior. Besides, the overall mean score of the teachers' self-evaluation surpassed the overall mean score of the students' evaluations of their instructors. One might wonder that instructors overestimated their capabilities, or that their students underestimated the instructors' qualifications, or probably both. The following table shows the Pearson correlation between the teachers' self-evaluation and their students' evaluations:

Table 4

Results of Pearson Correlation for the Relationship Between Teachers' Self-Evaluations and Their Students' Evaluations

Teachers \ Students		Evaluation Procedures	Teaching Method	Behavior	Students' Mean
Evaluation Procedures	Pearson Correlation	-.02	-.04	-.05	-.05
	Sig. (2-tailed)	.85	.69	.61	.67
	N	77	77	77	77
Teaching Method	Pearson Correlation	.03	.05	.02	.04
	Sig. (2-tailed)	.74	.62	.82	.72
	N	77	77	77	77
Behavior	Pearson Correlation	.09	.07	.07	.10
	Sig. (2-tailed)	.42	.51	.51	.37
	N	77	77	77	77
Cooperation	Pearson Correlation	.07	.10	.07	.07
	Sig. (2-tailed)	.54	.37	.54	.53
	N	77	77	77	77
Teachers' Mean	Pearson Correlation	.06	.03	.04	.05
	Sig. (2-tailed)	.56	.74	.72	.64
	N	77	77	77	77

The relationship between evaluation procedures (rated by the instructors) and evaluation procedures (rated by the students) was a very weak negative relationship. There was almost no relationship between the evaluation procedures as rated by the instructors and the students. As one might expect, this very weak relationship was not of statistical significance because the related *p* value appeared to be larger than the significance level. Similarly, the relationship between teaching method (rated by the instructors) and teaching method (rated by the students) was very weak. This very weak positive relationship did not reach statistical significance. As for behavioral aspect of the instructors, there was also a very weak positive relationship between the students' and the instructors' judgements. This weak relationship, unsurprisingly, did not reach statistical significance. Finally, there was a weak positive relationship between the overall instructors' evaluations and the

overall students' evaluations. This relationship, not unlike the previous ones, did not turn out to be statistically significant.

5. Discussion

The statistical analysis of the results indicated that there was no significant relationship between self-evaluation of the instructors and evaluation done by the EFL students, neither in terms of the overall evaluation, nor subcomponents of the evaluation. The results of the previous pertinent research indicated that the respondents prefer to under-report behaviors that might seem inappropriate or unfavorable by evaluators or other observers of their responses, and over-report behaviors viewed as appropriate and favorable (Dunning, Meyerowitz & Holzberg, 1989). One reason for this over rating on the part of teachers might lie in cultural issues.

Based on the results of many oriental culture studies, many Eastern cultures, including Iranian culture, promote a variety of 'selves' according to what is appropriate to a particular context (Heine, Lehman, Markus, & Kitayama, 1999). This is because many Eastern cultures consider the self as dependent on other people and the context in which it finds itself (Kitayama & Markus, 1999). In other words, the self is regarded as an extension of significant others including peers, friends and co-workers. This view of the self maintains that individuals are interconnected, and depends on each other for self-definition (Chu, 1985; Gao, 1998). Thus, an interdependent individual sees the self as less differentiated from others and is more likely to find ways to fit in with others.

On the other hand, a lot of students see the evaluation as a way of getting back at the instructor. There are students who have almost zero accountability and feel they are entitled to a good mark, regardless of whether they actually work. In their understanding, a good instructor is one who gives them high marks, has almost no expectations of their performance, does not give any homework, etc. If an instructor is strict, follows the rules and expects students to do the same, he or she almost receives bad scores on the evaluation. But it cannot be denied that students' ratings are vital and constructive. They provide feedback to the instructors which is important for their improvement. Students deserve a voice. It will allow them to have influence over the educational process. Based on the results of the investigation stated above, no correlation was found between the results of the instructors' self-ratings and students' ratings of their instructors. It seems that

both parties are not well-justified on characteristics and nature of the issue. Both instructors and students need to be more scientifically briefed on the issue to reduce bias and generate more trustworthy and down-to-the-earth results.

Moreover, the results of the study are in line with Shulman (1993) who accentuated the importance of discipline differences in the teaching evaluation. He proposed the discipline-specific evaluation of teaching as a reconnection between the evaluation and discipline. In this regard, clear differences were found in evaluating chemistry and history departments. Although, Mckechie (1996), after recounting the difficulties and pitfalls of comparing teaching in different departments, concluded that it is not a necessary and desirable effort.

As it was discussed in the previous studies such as Delvaux et al. (2013) and Richardson (2005), two main purposes are considered for the teacher performance evaluation. The first one is to determine the measure of teacher effectiveness. Then, administrative decision-making would be possible; it is also called quality assurance. The second one is to be able to provide diagnostic feedback for preparing a better teaching practice. In this way, teachers' strengths and weaknesses are identified and based on these identifications, they can improve their own teaching practices for further professional development. Consequently, the students' learning will enhance.

In this regard, it was concluded that at Isfahan Branch, although teachers' self-ratings and students' ratings are using for gathering evidence of teaching effectiveness, and the English Department wants to be able to make informed and objective decisions about retention, promotion, tenure, and pay raises but existence of such nonconformity between the results of rating scales is an obstacle in the way of achieving these goals.

But asking for evaluations regularly sends a clear message that teaching effectiveness matters, and not just in personnel decisions. Yet, probably the most important benefit of evaluations is the feedback the rating scales can provide directly to instructors, so that they can refine their courses and teaching practices to provide students with better learning experiences. By calling attention to teaching methods and outcomes, evaluations can play a positive role in improving the climate of teaching and learning at the university. But realization of these valuable outcomes depends on the university authorities' efforts in applying the achieved data in the post evaluation stages. Therefore, while implementing the teacher evaluation programs, it should be known that the outcomes obtained need to be

evaluated appropriately and the importance of this issue is not less than the validity and reliability of the instruments used in collecting the data for the evaluation.

6. Conclusion

Investigating the teacher evaluation takes a considerable significance and provides lots of implications because the education consumers are not by any means confined to the students. Actually, the main consumer of the education is the society which is going to benefit from the students' skills in the future. Therefore, present study bears a number of implications for educationalists, policy makers, and teacher trainers. Through conducting such a research, TEFL researchers, instructors, and language specialists will be provided with a clear picture of the current teacher evaluation scheme in the EFL context of academic settings. Therefore, it will help to direct the attention of the university authorities, supervisors, heads of departments, and EFL instructors themselves towards the importance and the impact of the teacher evaluation as an attempt to raise the quality of their performance and their students' outcomes. Moreover, it may enrich the literature in the field of performance evaluation of EFL teachers with the purpose of professional developments.

In addition, it may provide the higher education of Iran with helpful suggestions to improve the quality of the evaluation process in general at the universities especially in the English Departments and the information with which designing, implementing, and maintaining more effective EFL teacher evaluation practices could be possible. Another important implication of the study is related to those who are in charge of teacher education in the way that they are recommended to draw the attention of would-be teachers to the importance and usefulness of the teacher evaluation and make them aware of its crucial outcomes.

References

Ahmadi, M., & Sajjadi, S. (2009). Who should teach English for medical purposes (EMP). *Journal of Medical Education*, *13*(3), 135-140.

Angelo, T. A. (1996). Relating exemplary teaching to student learning. *New Directions for Teaching and Learning*, *1996*(65), 57-64.

Barton, H., & Shana, N. (2010). *Principals' perceptions of teacher evaluation practices in an urban school district*. California: University of the Pacific.

Bascia, N., & Jacka, N. (2001). Falling in and filling in: ESL teaching careers in changing times. *Journal of Educational Change*, *2*(4), 325-346.

Boggs, G. R. (1999). What the learning paradigm means for faculty. *American Association for Higher Education Bulletin*, *51(5)*, 3-5.

Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). Examining district guidance to schools on teacher evaluation policies in the Midwest Region. *Regional Educational Laboratory Midwest. Issues & Answers Report,* No. 030. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2007030.pdf

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work?. *Research in Higher Education*, *44*(5), 495-518.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. New Jersey: Association for Supervision and Curriculum Development.

Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, *36*, 1-11.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Sychology*, *57*(6), 1082-1090.

Ellett, C. D., Loup, K. S., Culross, R. R., McMullen, J. H., & Rugutt, J. K. (1997). Assessing enhancement of learning, personal learning environment, and student efficacy: Alternatives to traditional faculty evaluation in higher education. *Journal of Personnel Evaluation in Education*, *11*(2), 167-192.

Gao, G. (1998). An initial analysis of the effects of face and concern for other in Chinese interpersonal communication. *International Journal of Intercultural Relations*, *22*(4), 467-482.

Harmer, J. (2001). *The practice of English language teaching* (3rd ed.). London: Longman.

Harvey, J. N., & Barker, D. G. (1970). Student evaluation of teaching effectiveness. *Improving College and University Teaching*, *18*(4), 275-278.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard?. *Psychological Review*, *106*(4), 766.

Hinchey, P. H. (2010). Getting teacher assessment right: What policymakers can learn from research. *National Education Policy Center*. Retrieved from http://epicpolicy.org /publication/getting-teacher-assessment-right.

Isoré, M. (2009). Teacher evaluation: Current practices in OECD countries and a literature review. OECD education working papers. *OECD Publishing (NJ1)*. No. 23. Retrieved from http:// dx.doi.org/10.1787//223283631428.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *26*(1), 101-136.

Kennedy, M. (2005). *Inside teaching*. U.S.: Harvard University Press.

Kitayama, S., & Markus, H. (1999). The yin and yang of the Japanese self. *The coherence of personality*, 242-302.

McKeachie, W. J. (1996). Student ratings of teaching. *The professional evaluation of teaching*, Retrieved from http:// archives.acls.org/op/33_Professonal_Evaluation_of_Teaching.htm.

Mielke, P., & Frontier, T. (2012). Keeping improvement in mind. *Educational Leadership*, *70*(3), 10-13.

Palmer, B., & Marra, R. M. (2004). College student epistemological perspectives across knowledge domains: A proposed grounded theory. *Higher Education*, *47*(3), 311-335.

Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices (2ⁿᵈ ed.)* . Thousand Oaks, CA: Corwin Press.

Richards, J. C., & Rodgers, T. (2000). *Approaches and methods in language teaching*: *A description and analysis.* Cambridge: Cambridge University Press.

Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment and Evaluation in Higher Education*, *30*(4), 387-415.

Rigi, A. Ghaderi, M. Salimi, J. (2016). Comparing teachers' function evaluation by video data with other methods of teaching function evaluation, *Research in Curriculum Planning*. 13(23), 27-39.

Roessingh, H. (2006). The teacher is the key: Building trust in ESL high school programs. *Canadian Modern Language Review*, *62*(4), 563-590.

Scriven, D. Micael, C. Barbra, G. & Susan, T. (1987). *The evaluation of composition instruction* (2nd ed.). New York: Teacher's College Press.

Shulman, L. S. (1993). Forum: Teaching as community property: Putting an end to pedagogical solitude. *Change: The Magazine of Higher Learning*, *25*(6), 6-7.

Stake, R. (1987). The evaluation of teaching on campus. *Unpublished manuscript*, *University of Illinois at Urbana-Champaign*.

Theall, M., Abrami, P. C., & Mets, L. A. (2001). The student ratings debate: Are they valid? How can we best use them?. *New Directions for Institutional Research*, 109(1), 1-6.

Tudor, I. (1993). Teacher roles in the learner-centred classroom. *ELT journal*, *47*(1), 22-31.

Webb, L. D., & Norton, M. S. (2003). *Human resources administration: Personnel issues and needs in education*. Upper Saddle River, NJ: Prentice Hall.

Wren, S. (2006). Highly-qualified teachers are not necessarily high-quality teachers. Retrieved from http://www.balancedreading.com/teacherquality.html