

On the impact of teaching experience on EFL instructors' self-assessment of their instructional effectiveness

Pourjamal Ghouyjagh, Hakimeh¹

Sotoudehnama, Elaheh*²

Faghih, Esmail³

ID: IJEAP-1801-1152

Received: 17/04/2017 Accepted: 22/08/2017 Available online: 02/02/2018

Abstract

Teaching experience exerts a remarkable influence on instruction quality. The present study explores the role teaching experience plays in EFL university instructors' self-assessment (SA) accuracy and its congruence with students' assessment of it. Furthermore, it delves into the instructors' perceptions of how their SA of instructional effectiveness is altered as they obtain experience in their career. To this end, 106 EFL instructors' SA of their instructional effectiveness was compared to that of their 630 students. To this end, two versions of a questionnaire were administered to the instructors and their students. The findings suggest that the highly experienced (Hex, with over 21 years of experience) instructors' SA, unlike their relatively less experienced colleagues, i.e. less experienced (Lex, with 1-10 years of experience), and moderately experienced (Mex, with 11-20 years of experience), significantly diverges from their students' assessment. The statements of thirty-three interviewed instructors shed more light on the reasons behind such divergence besides the probable causes of instructional effectiveness erosion among Hex instructors. They blamed the students' and instructors' getting no education on accurate evaluation of teaching, adverse conditions dominating the higher education such as lack of standard hiring and evaluation system, job-burn-out and bias. The findings of this research can potentially contribute to EFL university instructor assessment, professional development, and education.

Keywords: self-assessment (SA), teaching experience, instructional effectiveness, student assessment of instructional effectiveness.

1. Introduction

Knowing oneself and awareness of one's abilities and inabilities can lead to self-improvement in every aspect of life. This issue gets utmost importance in assessing one's career performance and even greater importance when a career like teaching requires high mental, intellectual and physical abilities and a complex set of skills. Self-assessment (SA) is a strategy often mentioned in the literature on teacher professionalism, teacher

¹ English Language and Literature department, Faculty of Literature, Alzahra University, Tehran, Iran, Email: hkh_pourjamal@yahoo.com

² Corresponding Author, English language and literature department, Faculty of literature, Alzahra university, Tehran, Iran, Email: esotoude@alzahra.ac.ir

³ English language and literature department, Faculty of literature, Alzahra university, Tehran, Iran, Email: e_faghih@azad.ac.ir

empowerment, teacher evaluation, and professional development (Madsen, 2005). Yet, instructor SA is a “neglected strategy for facilitating professional growth” (Ross & Starling, 2005, p.146) particularly in higher education (Nilsson, 2012).

On both sides, student evaluations and instructor self-evaluations could be blurred by numerous factors such as inexperience and lack of training in teaching evaluation. It seems that university students’ end-of-semester ratings of instruction, which are usually carried out in Iranian higher education gives some hints regarding a number of restricted aspects of instruction. However, these ratings have mostly limited formative implications for instructors. Such a condition requires EFL instructors to acquire SA skill. In the present study, instructional effectiveness refers to the EFL university instructors’ judgments regarding their capabilities to teach EFL courses. Teacher’s SA of teaching effectiveness refers to teachers’ learning about students and themselves that comes from reflection on classroom experiences to make “judgments about the appropriateness or effectiveness of one’s own knowledge, performance, beliefs, products, or effects, so they can be improved” (Airasian & Gullickson, 1994, p. 6; italics added). SA undergoes changes during EFL instructors’ career span, a fact that brings instructional experience as a main actor into play. In the Iranian context, research on EFL university instructor SA is rare (Aghaei & Jadidi, 2013; Zarei & Afshari, 2012).

The problems addressed in this study are whether there are discrepancies between three experience groups of instructors’ SA and their students’ assessment of instruction effectiveness, and in what aspects these two evaluations converge or diverge. The questions and hypotheses we address in this study are:

1. Is there any statistically significant difference between Hex, Mex, and Lex EFL instructors in terms of the congruence between their SA and their students’ assessment of their instructional efficacy in general?
2. Is there any statistically significant difference between Hex, Mex, and Lex EFL instructors in terms of the congruence between their SA and the students’ assessment of their instructional efficacy in student engagement?
3. Is there any statistically significant difference between Hex, Mex, and Lex EFL instructors in terms of the congruence between their SA and the students’ assessment of their instructional efficacy in instructional strategies?
4. Is there any statistically significant difference between Hex, Mex, and Lex EFL instructors in terms of the congruence between their SA and the students’ assessment of their instructional efficacy in classroom management?

Hence, four null hypotheses related to the above-mentioned research questions were formulated.

2. Research background

Even when there is no external evaluation, a teacher's job can improve by timely and accurate use of SA as a metacognitive strategy. SA is of utmost importance in teachers' professional development. Regehr and Eva (2006) declared that in order to regulate one's competence, a teacher needs to self-assess gaps in his/her competence and also has to be willing to look for opportunities to close these gaps when identified.

Teachers use students' knowledge, feelings, attentiveness, body language, questions, facial expressions, opinions (Jones & Airasian, 1995), lesson content, students' actions and understandings, classroom management, curriculum, students' backgrounds, assignments, and school culture (Manouchehri, 2002) to evaluate their own instruction effectiveness. Generally, less experienced teachers are inclined to utilize student achievement scores as a measure of their instructional success (Madsen, 2005). Madsen (2005) concluded that novice teachers do not know what to look for to assess their teaching accurately. He added that only the most effective teachers used SA practices which were more likely to lead to positive changes. Rahimi, Ayati, and Asgari (2013) found meaningful relationship between teacher self-evaluation, and classroom management, and students' educational achievement.

The congruence between teachers' SA with an external evaluation such as students' evaluation of instructional effectiveness is an indicator of SA accuracy. Nevertheless, some studies showed that teacher self-ratings do not go with external evaluations; thus these studies did not consider SA as a dependable indicator of actual teaching effectiveness (Centra, 1979; Madsen, 2005). Syafar (2014) studied EFL teachers' SA of their competence to teach English and found that the participants' self-ratings did not match up with the results of a teacher competence test. He finally suggested that "self-assessment should be endorsed with teaching performance assessment to have more reliable data for validating EFL teachers' self-rating abilities in English teaching practices" (p.304).

Teachers are different from each other since the experiences they have attained throughout their teaching career are different (Zarei & Afshari, 2012). Typically, people learn from their experiences. Madsen (2005) described Dewey's (1916) philosophy in this regard as follows:

What is crucial in gaining experience through reflection is the intentional effort to identify a problem, collect evidence and information to solve the stated problem, and systematically test proposed ideas - thus the surgery of thinking about doing and doing with thinking. By combining thinking and doing into a synergistic method of inquiry, teachers' knowledge and practice of classroom teaching and learning can advance (p. 17).

Since there is no solid experience criterion for teaching, researchers have taken diverse experience span of instruction as the basis for their studies. Nevertheless, literature (Berliner, 1987; Ferry & Ross-Gordon, 1998) indicates that years of teaching cannot solely

be the indicator of expertise in teaching. Therefore, how one utilizes experience is the more crucial factor in self-improvement.

Tschannen-Moran and Woolfolk-Hoy (2007) investigated the association between teaching self-efficacy, and teacher experience. They found that experienced teachers' general teaching self-efficacy was significantly higher than that of novice teachers. Nevertheless, there was no difference in teaching self-efficacy and in student engagement. Tschannen-Moran and Woolfolk-Hoy (1998) indicated that "change is difficult. Even when changes are made for the better, they are uncomfortable and stressful" (p. 236). It gets more difficult with time, not because the learning is more difficult (although this may also be true), rather "because the older we get the less energy we are willing to exert to learn something new (and the fewer the number of people who have the authority to tell us we must learn it)" (Regehr & Eva, 2006, p. 36).

When novice teachers start their career, they recurrently stumble upon a reality shock as they face up to the difficulty of the teaching undertaking (Tschannen-Moran et al., 1998). "A 'get tough' attitude may result for those teachers who conclude that the constraints of teaching are formidable and that the resources for dealing with the problems are weak" (Tschannen-Moran et al., 1998, p. 232).

Hogan, Rabinowitz, and Craven (2003) stated that experienced teachers' information store and teaching and learning skills network are more integrated and organized than those of novice teachers. Student achievement is very important for expert teachers, while novice teachers pay more attention to class interest (Hogan et al., 2003). In addition, experienced teachers are more evaluative of teaching situations than novice ones (Sabers, Cushing, & Berliner, 1991). Hogan et al. (2003) also indicated that novice teachers are unable to monitor and accurately evaluate teaching events.

Thus far, a comparison of SA of instructional effectiveness of EFL university instructors and the assessment of their students of it considering the role experience might play in it have not been investigated.

3. Methodology

3.1. Participants

Considering the objective of the research two groups of participants were recruited. The first group, EFL instructors, participated in both quantitative and qualitative phases of the research by filling out a revised version of Teacher Sense of Efficacy Scale (TSES) devised by Tschannen-Moran and Woolfolk-Hoy (2001) (Appendix 1), as well as by subsequently responding to the interview questions (Appendix 3). The second group of participants, the above-mentioned instructors' students, participated in the quantitative phase and filled out the student version of TSES (Appendix 2) which was an altered form of the original questionnaire modified for the students.

The EFL instructors who participated in this research taught EFL courses including both general English and content courses in the fields of Teaching English as a foreign language, English literature, English translation, and linguistics. The selection of the participants was based on convenience sampling.

One hundred and six instructors filled out the questionnaire and 33 of them were interviewed in order to explore the role of experience in EFL instructors' SA besides the reasons why matches and mismatches exist. In terms of experience, the instructors included 33Hex, with over 20 years of experience in university teaching in various fields of EFL, 37 Mex, with 11-20 years of experience, and finally 36 Lex instructors, with 1-10 years of experience. Since a university instructor in Iran teaches for 30 years, the researchers assumed each decade of their career as a time span in this study. The instructors of each group were mostly chosen based on their accessibility. Thirty-three instructors including 12 Lex, 9 Mex, and 12 Hex instructors were also interviewed based on convenience sampling. Six hundred and thirty students were selected randomly from different current classes of the same instructors who participated in this study. Approximately six students filled out the student version of the questionnaire to evaluate each instructor. The sampling procedure implemented in this phase was random sampling.

3.2. Instrumentation

TSES comprises three sub-scales to test teachers' sense of efficacy: (1) efficacy in student engagement, (2) efficacy in instructional strategies, and (3) efficacy in classroom management. The reliability of both instructor and student versions of TSES which were calculated through Alpha Cronbach are presented in table 3.1. For reliability assurance, both instructor and student versions were piloted with 40 participants for each group before large scale administration of the instrument. It is worth noting that based on Dornyei (2007), reliability coefficients of over .70 is acceptable in L2 research.

Table 1: *Reliability Analysis of TSES Instructor and Student Versions*

| | Instructor version | Student version |
|-------------|--------------------|-----------------|
| General | .91 | .96 |
| Engagement | .77 | .91 |
| Instruction | .83 | .91 |
| Management | .77 | .90 |

The questionnaires were distributed among the EFL students and instructors of 12 universities including Allameh Tabatabaei University, Alzahra University, University of Guilan, Kharazmi University, Nabye Akram Non-profit University (Tabriz), Qom University, Rasht Payame Noor University, Semanan University, Shahid Madani Azarbayjan University, Tabriz University, Tabriz Azad University, and finally Tehran University. To answer the quantitative questions, the three experience groups of instructors'

self-rating averages were compared to those of their students' ratings. The instrument implemented in the qualitative phase of this research was a researcher-developed semi-structured interview with seven questions (appendix 3). All interviews were audio recorded. The interviewees were made assured as to the confidentiality and anonymity of the interviewed content before and during the interviews.

3.3. Research Design

A mixed methods design was used in this study. As the major goal of the present study is comparing two sets of ratings by instructors and students, a survey design was opted for because surveys can elicit comparable information from respondents (Mackey & Gass, 2005). Moreover, Dornyei (2003) considered surveys the most appropriate research design for teacher self-evaluation as well as teacher evaluation (done by students in this study). The design of the qualitative part of this study was based on thematic analysis.

4. Results and Discussion

The present research intended to explore the role of teaching experience in correspondence between EFL instructors' SA and their students' assessment of general teaching efficacy and its sub-components namely student engagement, instructional strategies, and classroom management. To this end, the assessment made by the three groups of experience i.e., highly experienced (Hex), moderately experienced (Mex), and less experienced (Lex) were compared with those of their students. In order to compare Lex, Mex, and Hex instructors' SA with their students' assessment in general teaching effectiveness, three Mann-Whitney U tests were run. Besides, to compare the three groups of instructors with their students in instructional strategies, classroom management, and student engagement, nine further tests were conducted. Since the same test (Mann-Whitney U test) was run for 12 times on the same sample which might have increased running the risk of being trapped in type I error, Bon-Ferroni correction for α (Frane, 2015) was used to avoid the risk. In the correction for α , and in order to control for "multiple comparisons" (Dancey & Reidy, 2011, p. 308), the intended significance level (i.e., .05) was divided by the number of tests (i.e., 12) which gave a significance level of .004. Therefore, .004 was the basis for reporting all of the results of the quantitative part of the present research. The discussion of effect sizes is based on Rosenthal (1994).

Table 2 presents the results of Mann-Whitney U test run to evaluate the significance value of the differences observed in the mean ranks. Based on the table, the difference between Lex instructors and their students in assessing these instructors' general teaching effectiveness is statistically insignificant as the related p-value indicates; moreover, the difference is small as the related effect size indicates ($U = 2078.5$, $Z = -2.043$, $p > .004$, $r = .15$). The difference between Mex instructors and their students in assessing these instructors' general teaching effectiveness is statistically insignificant as the related p-value indicates; furthermore, the difference is small as the related effect size indicates ($U = 2040$,

$Z = -2.478$, $p > .004$, $r = .17$). However, the difference between Hex instructors and their students in assessing these instructors' general teaching effectiveness is statistically significant as the related p-value indicates; besides, the difference is medium as the related effect size shows ($U = 1568.5$, $Z = -4.553$, $p < .004$, $r = 0.3$). As the p-values and the effect sizes of the three groups indicate, the first null hypothesis is rejected in case of Hex instructors but not rejected in case of Mex and Lex instructors. This means that experience plays a role in the in/congruence between the EFL instructors' SA and their students' assessment of general instructional efficacy.

Table 2: *Mann-Whitney U Test between Two Groups of Instructors and Their Students' Assessment in Terms of Instructors' General Instructional Efficacy*

| General Instructional Efficacy | Mann-Whitney U | Wilcoxon W | Z | Asymp. Sig. (2-tailed) | r |
|--------------------------------|----------------|------------|--------|------------------------|------|
| Lex | 2078.500 | 13104.500 | -2.043 | .041 | 0.15 |
| Mex | 2040.000 | 17616.000 | -2.478 | .013 | 0.17 |
| Hex | 1568.500 | 19523.500 | -4.553 | .000 | 0.3 |

To answer the second research question, the Lex, Mex, and Hex instructors' SAs were compared to the assessment of their students with respect to their efficacy for student engagement. Table 3 reveals the results of Mann-Whitney U test which was run to evaluate the significance value of the differences observed in the mean ranks. The second null hypothesis is rejected only in case of the Hex instructors. Based on table 3, the difference between Lex instructors and their students in assessing these instructors' student engagement effectiveness is statistically insignificant as the related p-value indicates; in addition, the difference is small as the related effect size shows ($U = 2139.5$, $Z = -1.832$, $p > .004$, $r = .13$). The difference between Mex instructors and their students in assessing these instructors' student engagement effectiveness is statistically insignificant as the related p-value indicates; furthermore, the difference is small as the associated effect size signifies ($U = 2218.5$, $Z = -1.91$, $P > .004$, $r = .13$). The difference between Hex instructors and their students in assessing these instructors' student engagement effectiveness is statistically significant as the related p-value indicates; besides, the difference is almost medium as the related effect size points to ($U = 1779$, $Z = -3.937$, $p < .004$, $r = .26$). The p-values and the effect sizes of the three experience groups indicate that there is difference among the Lex, Mex, with Hex instructors in their SA's

correspondence with their students' assessment of these three groups' student engagement effectiveness.

Table 3: Mann-Whitney U Test between Two Groups of Instructors and Their Students' Assessment in Terms of Three Groups of Instructors' Student Engagement Efficacy

| Student Engagement | Mann-Whitney U | Wilcoxon W | Z | Asymp. Sig. (2-tailed) | r |
|--------------------|----------------|------------|--------|------------------------|------|
| Lex | 2139.500 | 13165.500 | -1.832 | .067 | 0.13 |
| Mex | 2218.500 | 17794.500 | -1.910 | .056 | 0.13 |
| Hex | 1779.000 | 19734.000 | -3.937 | .000 | 0.26 |

Table 4 presents the results of Mann-Whitney U test for comparing the SA of instructors to their students' assessment of student engagement efficacy. The third null hypothesis is accepted in case of the Lex and Mex instructors and rejected in case of the Hex ones. The table shows that the difference between Lex instructors and their students in assessing these instructors' instructional strategies effectiveness is statistically insignificant as the related p-value indicates; in addition, the difference is small as the associated effect size signifies ($U = 1979.5$, $Z = -2.391$, $p > .004$, $r = 0.18$). The difference between Mex instructors and their students in assessing these instructors' instructional strategies effectiveness is statistically insignificant as the related p-value indicates; and the difference is also small as the related effect size shows ($U = 1897.5$, $Z = -2.936$, $p > .004$, $r = .2$). Finally, the difference between Hex instructors and their students in assessing these instructors' instructional strategies effectiveness is statistically significant as the related p-value indicates; moreover, the difference is medium as the related effect size indicates ($U = 1372$, $Z = -5.135$, $p < .004$, $r = .34$). Thus Lex and Mex instructors are different from their Hex colleagues in this respect.

Table 4: Mann-Whitney U Test Between Two Groups of Instructors and Their Students' Assessment in Terms of Three Experience Groups of Instructors' Instructional Strategies Efficacy

| Instructional Strategies | Mann-Whitney U | Wilcoxon W | Z | Asymp. Sig. (2-tailed) | r |
|--------------------------|----------------|------------|--------|------------------------|------|
| Lex | 1979.500 | 13005.500 | -2.391 | .017 | 0.18 |
| Mex | 1897.500 | 17473.500 | -2.936 | .003 | 0.2 |
| Hex | 1372.000 | 19327.000 | -5.135 | .000 | 0.34 |

As the results indicate, the third null hypothesis is accepted in case of the Lex instructors and rejected in case of the Hex and Mex instructors. This means that experience plays a role in in/congruence between the instructors' SA and their students' assessment of instructional strategies.

The last research question addressed the Lex, Mex, and Hex instructors' SA of their own classroom management and these instructors' students' assessment of it.

As table 5 reveals, the null hypothesis that there is no statistically significant difference between the Lex, Mex, and Hex instructors' SAs and their go-togetherness with the students' assessment of classroom management effectiveness, is verified in case of Lex and Mex instructors and rejected in case of the Hex instructors. Mann-Whitney U test results indicate the difference between Lex instructors and their students in assessing these instructors' classroom management effectiveness is statistically insignificant as the related p-value indicates; and this difference is small as the associated effect size signifies ($U = 2263$, $Z = -1.401$, $p > .004$, $r = .1$). The difference between Mex instructors and their students in assessing these instructors' classroom management effectiveness is statistically insignificant as the related p-value indicates; furthermore, the difference is small as the related effect size shows ($U = 2178$, $Z = -2.039$, $p > .004$, $r = .14$). Finally, the difference between Hex instructors and their students in assessing these instructors' classroom management effectiveness is statistically significant as the related p-value indicates; and the difference is approximately medium as it can be noticed from the related effect size ($U = 1785.5$, $Z = -3.919$, $p < .004$, $r = .26$).

Table 5: Mann-Whitney U Test between Two Groups of Instructors and Their Students' Assessment in Terms of Three Experience Groups of Instructors' Classroom Management Efficacy

| Classroom management | Mann-Whitney U | Wilcoxon W | Z | Asymp. Sig. (2-tailed) | r |
|----------------------|----------------|------------|--------|------------------------|------|
| Lex | 2263.000 | 13289.000 | -1.401 | .161 | 0.1 |
| Mex | 2178.000 | 17754.000 | -2.039 | .041 | 0.14 |
| Hex | 1785.500 | 19740.500 | -3.919 | .000 | 0.26 |

The above-mentioned results lead us to the conclusion that experience plays a role in this case, too. Thus, as mentioned above, the null hypothesis that there is no statistically significant difference between the Lex, Mex, and Hex instructors' SAs and their correspondence with the students assessment of classroom management effectiveness, is verified in case of Lex and Mex instructors and rejected in case of the Hex instructors.

As the results of the study indicate, as the EFL instructors' teaching experience increases the go-togetherness of their SA and their students' assessment decreases. Both Lex and Mex instructors' SAs go well with their students' assessment of instruction, but the Hex instructors' SA does not agree with their students' assessment. In other words, the instructors who are spending the last 10 years of their career do not see their instruction as their students see it. This significant difference is observed in the SA and assessment of instructional efficacy in general and its three subcomponents namely, student engagement efficacy, instructional strategies efficacy, and classroom management efficacy.

Although research findings show that experience plays a remarkable role in raising teachers' instructional efficacy (Fakhary, 2014; Woolfolk-Hoy & Burke-Spero, 2005), and although Pajares (1997) claimed that through experience people enjoy the opportunity to evaluate and consider the results of their own actions, the findings of this research suggest that experience especially in its high levels plays an adverse role in the congruence between the Iranian Hex EFL instructors' SA and their students' assessment of instruction. It is evident that as the experience level of EFL instructors increases, their perception of their own teaching effectiveness becomes more distant from the perception of their students of it. The results indicate that SA of the Lex and Mex instructors are closer to the assessment of their students. If we consider this kind of agreement as one of the indicators of accuracy of SA, the findings of this research goes contrary to Madsen's (2005) that Lex teachers do not know what to look for to self-assess their teaching accurately.

Khodaverdi (2009), and Rastegar and Memarpour's (2009) studies revealed no significant relationship between the teachers' sense of efficacy and their experience. A further study found that instructors with more teaching experience and higher levels of education had higher levels of teaching efficacy (Hoy & Woolfolk, 1993). In CapaAydin and Woolfolk-Hoy's (2005) study, pre-service teachers with more teaching experience tended to enjoy less sense of teaching effectiveness.

For Hex and Lex, experience is essential for developing SA ability. The Hex claim that they have gained this ability through experience, yet, ironically, their SA does not match with the assessment of their students based on the findings of the quantitative part of the present research. The Lex are hopeful to acquire SA skill as their teaching experience increases. Nevertheless, Hex and Mex instructors protest about the existence of inconvenient conditions for comprehensive and accurate SA.

In line with the findings of the present study, Gurvitch and Metzler (2009) found that there is positive association between Lex teachers' SA of efficacy and their experience of teaching. Hogan et al. (2003) found that while student achievement is very important for Hex instructors, Lex instructors pay more attention to class interest. In the qualitative section, we also found that student engagement was more a concern for Lex and Mex instructors than for Hex ones. Experience proved to be a constructive factor in developing instructional strategies, classroom management (Loreman, Sharma, & Forlin, 2013) and student engagement (Akbari & Moradkhani, 2010). Corroborating the quantitative findings of this research, student engagement was more of concern for the Mex and Lex than for Hex instructors in the qualitative part. Generally, student engagement is of utmost importance for Iranian EFL instructors since they believed that students are less motivated nowadays due to unemployment problems they anticipate to face after graduation.

However, according to Choy, Wong, Lim, and Chong (2013), instructional strategies and classroom management are more challenging areas for less experienced teachers.

Tshannen-Moran and Woolfolk Hoy (2002) did not find any significant difference between novice and experienced teachers' efficacy for student engagement which goes contrary to Fakhary (2014) whose study demonstrated that experienced English teachers had higher efficacy for student engagement than the novice ones. However, classroom management concerned the EFL instructors in the present research far less than other aspects of teaching based on the interviews.

Experienced instructors are more evaluative of teaching situation than less experienced ones (Sabers et al., 1991). Hogan et al. (2003) also indicated that novice teachers are unable to monitor and accurately evaluate the teaching events. The findings of the present study do not prove these findings as Lex instructors in the current study appear to be highly evaluative of their instruction. The Mex instructors stood in between but much closer to Lex instructors in their SA's congruence with their students' assessment.

Tschannen-Moran and Woolfolk-Hoy (1998) asserted that change is difficult. Regehr and Eva (2006) believed that the older people get, the less energy they exert to learn new things and there is a fewer number of people who have the authority to tell them they must learn something. On the other hand, Lex or even Mex instructors may have higher motivation to change for better because of being more energetic and having less stable job status. Generation gap with students is also less in these two groups' cases. Besides, Lex instructors are more in need of surviving in the academic environment.

The probable reasons of lack of motivation to teach effectively or instructional effectiveness erosion among some mostly Hex instructors were also explored. The instructors referred to several external and internal variables which are intricately interconnected. It is interesting that these factors are mainly stated by Hex instructors themselves which might be indicative of the fact that they are more familiar with the phenomenon. The interviewees from all three participant groups referred to some external discouraging factors leading to instructional effectiveness erosion including:

- 1) The adverse conditions of educational context,
- 2) Not being acknowledged as a qualified instructor by the system,
- 3) Unwelcome financial conditions and disparity between instructors' workload and payment,
- 4) Lack of motivation among students, and their being after degree and score rather than learning,
- 5) Instructors' especially Hex ones' enjoying a permanent employment status, and their not being respondent to any higher level authorities.
- 6) Lack of constant standard evaluation of instructors' teaching quality.

The internal discouraging factors include:

- 1) Job burn-out,
- 2) Being aged and losing energy by time,

- 3) Being entangled in repetition,
- 4) Not being interested in one's job or specialist field (EFL),
- 5) Lack of motivation,
- 6) Not having constant SA,
- 7) Not being conscientious and committed enough,
- 8) Overconfidence

In accordance with these findings, Furnham and Chamorro-Premuzic (2005) found that individual personality characteristics can influence the evaluation of one's own self. In addition, Sargeant, et al.(2010) considered personal attributes as important in the whole process of SA. They referred to attributes and emotions such as motivation, confidence, curiosity, engagement, mindfulness, and self-directedness.

Although Zarei and Afshari (2012) found no relationship between experience level of their participants and their motivation level, the current study's findings revealed a mismatch between EFL instructors' experience level and their motivation to keep teaching effective. Dunning and Helzer (2014) also indicated that overconfidence undermines the accuracy of SA.

By reforming some policies in university level such as encouraging the different experience groups based on their needs, the conditions might change for better. TLRP (2006) also suggested that strategies for maintaining commitment in initial and enduring professional development should be devised to distinguish between the needs of teachers in different stages of their professional lives.

Experienced teachers' efficacy beliefs seem to be dead set against change once established (Tschannen-Moran & Woolfolk-Hoy, 2007). It gets more difficult with time, simply not because the learning is more difficult (although this may also be true), rather "because the older we get the less energy we are willing to exert to learn something new (and the fewer the number of people who have the authority to tell us we must learn it" (Regehr & Eva, 2006, p.36).

Teachers do not necessarily maintain their effectiveness over time (TLRP, 2006). "Teachers in later years are at greater risk of becoming less effective though these are still a minority" (TLRP, 2006). CapaAydin and Woolfolk-Hoy (2005) also found the same results with experienced teachers.

Based on the interviews, EFL instructors are pessimistic about the feedback they receive from university administrators, colleagues, and students as three main sources of data to evaluate and improve their instruction. The important point is that university authorities could provide other sources of information such as supervisors, consultants, and colleagues in a systematic and formative way to develop SA skills. Yet, all they do is delivering summative results of end-of-semester ratings to instructors. This is while the importance of

colleagues' and supervisors' continuous assistance is emphasized in the literature (Sargeant et al., 2010).

The most important and often the only source of feedback for the evaluation of instruction for university instructors is students' end-of-semester ratings of their instructors' teaching. Yet, a majority of the EFL instructors who participated in this research do not consider these ratings as dependable for reasons such as their summative and quantitative nature, the instruments' defective structure and questions, students' being disqualified to assess instructors' work, students' considering unrelated factors rather than teaching quality, instructors' popularity, instructors' score assignment leniency, or even fear of instructors' retaliation in case he is not rated highly by the students. Aleamoni's (1999) participants, in accord with the findings of this study, also expressed their concerns as to the student ratings' being mainly a popularity competition among instructors. Aleamoni's (1999) participants consider student evaluation forms as both unreliable and invalid. They think that the evaluations are not formative in nature; in addition, the instructors are left alone with the summative score-based results and there is no professional help from university administrators at all. In fact, "formative feedback should be non-evaluative, supportive, timely, and specific" (Shute, 2008, p. 153).

As student ratings are the only source of external data for SA and instruction improvement in Iranian University context, they should be taken into account, used, and complemented by other SA strategies and sources to make logical changes and improvements. Worrell and Kuterbach (2001) have warned against student ratings as being stand-alone evaluation measures, as students are not usually qualified to rate teachers on curriculum, classroom management, content knowledge, collegiality, or other aspects of effective teaching.

Sargeant et al. (2010) maintained that it is imperative that SA be used systematically and involve discussion with others, such as a mentor or professional consultant in order for SA to improve performance. The participants of the present research also mostly blame the external conditions if they do not improve based on SA. Eva and Regehr (2008) emphasized the role of professional help following SA and declared that personal, unguided SAs merely do not lead to performance improvement sufficiently.

In general, the participants of the present study from all groups claimed that the reasons they might not be competent enough to evaluate their teaching might be not being trained for this purpose neither during their education, nor after starting their career, as well as not having enough time and energy to do so due to being busy teaching and doing research. In line with these findings, Lanyon and Hubball (2008) accentuated that teachers need to receive education to have accurate SA.

Elliot (1995) also found that neither training, and experience, nor professional culture had allowed teachers to develop the ability required to become reflexive, self-aware and therefore able to self-assess. He asserted that teachers are methodologically adrift, meaning

that they are uncertain of what questions to ask of themselves, what kinds of data to collect by what methods and how to evaluate it when it has been collected. The present research also revealed that generally the EFL instructors do not know how to evaluate their job in that they do not know what questions to ask of themselves or what instrument to use to evaluate their own instruction and what to do with the raw data they obtain after evaluation. On the other hand, novice teachers might be cognitively overwhelmed during their first experiences of teaching due to high performance demands and thus could benefit from supportive feedback to decrease the cognitive load (Shute, 2008).

Bias in SA is inevitable and people are usually unconscious of the biases in their SA (Zell & Krizan, 2014). The findings of the present research indicate that SA of instructional effectiveness is more subjective than objective because it is mostly done based on feelings rather than the facts received from outside sources. To solve the problem of bias to some extent, three solutions were recommended by the participants of this study. SA needs to be:

- 1- Ongoing and triangulated by collecting data from other sources than only students' comments and instructors' feelings,
- 2- Based on the reactions of a large number of students,
- 3- Based on the alumni's feedback.

In line with the last suggestion of the instructors to make student SAs more reliable, Aleamoni's (1999) instructor participants also recommended that students should graduate from the course and the university for quite a few years to make an accurate judgment about teaching effectiveness. In accord with what the participants of the present study suggested, Little, Goe, and Bell (2009) proposed that the data for SA should be gleaned from various rather than a single source. In addition, Koziol and Burns (1987) found that the accuracy of teachers' SA increases when the process is repeated and the results are compared.

Hex instructors refer to their own internal feelings and experience to realize the incentive behind the feedback, including admiration and criticism, they obtain from their students for SA. Lex instructors have more constant and triangulated SA in order to increase its accuracy. These instructors prefer to utilize the feedback they get from a higher number of students to make their SA dependable. These two findings can suggest that Lex instructors are more concerned about their own SA accuracy that goes with our quantitative findings.

Eva and Regehr (2008) found that bias is more likely to contaminate people's SA. They referred to memory bias, information disregard, trying to sustain an optimistic attitude and insufficient feedback as the cases of inaccuracy of SA. Even though self-perceptions are often contaminated by bias and the filter of the self, "Reconciliation of feedback with self-perceptions, in fact, appeared to be a primary step and fundamental to assimilating, accepting and using external feedback" (Sargeant, 2008, p. 50).

Lex feel the need for being professionally helped to do accurate SA, yet, Hex instructors do not feel the need for such assistance at all. This might indicate that the Hex feel overconfident in that they are competent self-assessors by themselves, the finding which goes contrary to the findings of the quantitative section. Cardenas' (2009) study on meta-cognitive awareness-raising training among Chilean EFL teacher participants led to improvements in proficiency and general awareness, and also resulted in remarkable changes in less experienced teachers. Her finding corroborates that of the present study in that the Lex instructors see themselves more in need of help to become proficient self-assessors.

Corresponding to the above-mentioned statements, Lanyon and Hubball (2008) referred to teachers' need to receive education to enhance assessment concerning self. Moreover, Kuiper and Pesut (2004), and Westburg and Jason (2001) emphasized the higher education instructors' need to be formally trained on SA skills. In addition, Dunning and Helzer (2014) argued that teaching people how to make better judgments, results in more accurate SA. McNamara and O'Hara (2008) also considered systematic training of SA in extended period of time as vital for developing this skill. Therefore, large-scale policies are blamable for the possible instruction inefficacy or inaccurate SA among EFL instructors.

The interviewees were asked about the importance of the three dimensions of instructional efficacy in their SA. Student engagement is the first priority for the EFL instructors in Iranian universities. This finding verifies the EFL instructors' complaints as to their students' loss of motivation due to disappointing unemployment conditions after graduation.

It appears that experience plays a role in giving priority to particular aspects of teaching. The Mex, and then Lex care more about keeping students engaged and motivated in order to keep their teaching quality. Manouchehri's (2002) preservice teacher participants focused on self, lesson content, the students' actions and understandings, the activity or task of the lesson, classroom management and control, the curriculum, the teachers' actions, the students' backgrounds, assignments, and school culture. Loreman et al. (2013) found that teachers' experience developed their instructional strategies and classroom management and Akbari and Moradkhani (2010) concluded that it enhances student engagement.

Lex instructors care more about their teaching strategies and knowledge of the field in the SA of instruction. Mex instructors consider all three dimensions of teaching, i.e., teaching strategies, student engagement, and classroom management as equally important in SA. However, the Mex instructors do not see their teaching strategies as their students see it. Classroom management was no concern for the studied groups in the qualitative section, yet the findings of the quantitative section indicate that Hex in structures are not successful self-assessors of their classroom management and their evaluations, in this aspect, were significantly divergent from their students' (those who are managed) assessment. Loremen

et al.'s (2013) finding goes contrary to the finding of the present study in that experience was a constructive factor in developing instructional strategies and classroom management. As to the reasons of possible mismatch between the all participant instructors' and their students' assessment, the instructors refer to two main reasons of mismatch between their SA and the assessment of their students from their work: 1) the score that students assign influences their assessment of the instructors, 2) instructors are unable to evaluate their work accurately. Besides several other reasons are listed:

- 1) Students and instructors are not trained to have accurate evaluations of teaching,
- 2) Students might not have insight into the rationale behind what their instructors do,
- 3) There is no constant and reliable communication and mutual understanding between instructors and students,
- 4) Two or three extreme scores in end-of-semester ratings of the instructors' teaching might impact the mean score greatly,
- 5) Some instructors may be influenced by the appreciation of some students who are extraordinarily absorbed to them, and shape an inaccurate positive portrait of themselves,
- 6) Some instructors' and their students' age distance may even be half a century,
- 7) Instructors and students have different back grounds; thus they might see the same phenomenon differently,
- 8) Students compare their instructors while instructors might never know how his/her colleague is performing,
- 9) Some instructors develop too much friendly relationship with students and the ones who do not do so might be under-rated by students,
- 10) Humans are self-centered and biased creatures, a fact which leads to over-assessment of one's performance,
- 11) Instructors do not know about their students' expectations (maybe due to lack of communication),
- 12) Context variation and also variation across majors create divergent perceptions.

To solve the problem of mismatch between instructors' SA and their students' assessment, and to solve the third above-mentioned problem, Kern (1995) referred to the necessity of open communication between students and teachers vis-à-vis teaching and learning beliefs and practices as a major implication of his study. Kern also affirmed that "by listening closely to our students, by identifying mismatches in beliefs, and by clearly explaining why we do what we do in the classroom, it may be possible to significantly allay student frustration" (p.81). Such constructive discussions exist in many European universities between faculty members and their students in personal face-to-face meetings called 'development sessions'.

EFL instructors from all three groups blame the students as the source of divergence between their SA and the students' assessment. However, Lex instructors blame the

instructors' own inaccuracy in their SA as the cause of divergence between the assessments of the two groups. Lex instructors thought that the reason might be inexperience or bias in SA from the side of the instructors.

Mostly Mex and leastly Lex instructors think that their SA and their students' assessment match. This is contrary to the findings of the quantitative section of the present research which shows that the Lexs' SA has the highest congruence with their students' assessment.

5. Conclusion

This research suggests that as EFL instructors become more experienced, their perception and assessment of their own instructional efficacy become far-off from what their students evaluate although experience proves to help highly experienced instructors to use consultation with colleagues to improve their instructional effectiveness. Therefore, experience is a main factor which determines the instructor-student assessment correspondence. The findings of the qualitative phase were illuminating in helping us understand the "why" and "how" of the impact of experience on teaching effectiveness, instruction effectiveness SA by EFL instructors and instructor-student assessment in/congruence. The interviews revealed more facts about the causes of such divergence besides the probable disincentives leading to instructional effectiveness erosion among Hex instructors. The instructors hold the students' and instructors' getting no education on accurate evaluation of effective teaching, unfavorable circumstances dominating the higher education system for instance lack of standard instructor hiring and evaluation criteria, and job burn-out or even bias as the major undesirable elements in this regard.

References

- Airasian, W., & Gullickson, A. (1994). Teacher self-assessment: Potential and barriers. *Kappa Delta Pi Record*, 30, 6-10.
- Aghaei, P., & Jadidi, E. (2013). The effect of EFL teachers' language awareness and gender on their reflectivity. *International Journal of Language Learning and Applied Linguistics World*, 4(3), 94-104.
- Akbari, R., & Moradkhani, S. (2010). Iranian English teachers' self-efficacy: Do academic degree and experience make a difference? *Pazhuhesh-e Zabanha-ye Khareji*, 56, 25-47.
- Aleamoni, M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13(2), 153-166.
- Berliner, C. (1987). Simple views of effective teaching and a simple theory of classroom instruction. In C. Berliner & B. Rosenshine (Eds.), *Talks to Teachers* (pp.93-110). New York: Random House.

- CapaAydin, Y., & Woolfolk-Hoy, A. (2005). What predicts student teacher self-efficacy? *Academic Exchange Quarterly*, 9, 123-128.
- Cardenas, V. (2009). *Meta-cognitive awareness-raising in EFL teacher education with special reference to Chilean context*. Unpublished Phd dissertation. Birmingham University.
- Centra, J. (1979). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Choy, D., Wong, A. F. L., Lim, K. M., & Chong, S. (2013). Beginning teachers' perceptions of their pedagogical knowledge and skills in teaching: A three year study. *Australian Journal of Teacher Education*, 38(5), 68-79.
- Dancey, C. P. & Reidy, J. (2011). *Statistics without math for psychology*. Essex: Pearson Education.
- Dornyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, N.J.: Lawrence Erlbaum.
- Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dunning, D., & Helzer, E. G. (2014). Beyond the correlation coefficient in studies of self-assessment accuracy: Commentary on Zell & Krizan (2014). *Perspectives on Psychological Science*, 9(2), 126-130.
- Eva, K. & Regehr, G. (2008). "I will never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*. 28(1),14-19.
- Fakhary, M. (2014). *The impact of MA TEFL courses in Iran on prospective teachers' self-efficacy* (Unpublished MA thesis), Alzahra University, Iran, Tehran.
- Ferry, M. & Ross-Gordon, M. (1998). An inquiry into Schon's epistemology of practice: Exploring links between experience and reflective practice. *Adult Education Quarterly* 48, 98-112.
- Frane, A. (2015). Are per-family Type I error rates relevant in social and behavioral science? *Journal of Modern Applied Statistical Methods*, 14(1),12-23.
- Furnham, A. & Chamorro-Premuzic, T. (2005). Estimating one's own and one's relatives' multiple intelligence: A study from Argentina. *Spanish Journal of Psychology*, 8(1), 12-20.
- Gurvitch, R., & Metzler, M. (2009). Teaching efficacy in preservice physical education teachers. *Teaching and Teacher Education*, 25, 437-443.

- Hogan, T., Rabinowitz, M., & Craven, J. (2003). Representation in Teaching: Inferences from research of expert and novice teachers. *Educational Psychologist*, 38(4), 235-247.
- Hoy, W., & Woolfolk, A. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal*, 93(4), 355-372.
- Jones, A., & Airasian, P. (1995). Through their eyes: Teacher self-assessment. *Mid-Western Educational Researcher*, 8(4), 12-16.
- Kern, R. (1995). Students' and teachers' beliefs about language learning. *Foreign Language Annals*, 28(1), 71-92.
- Khodaverdi, N. (2009). *The relationship between teachers' and learners' expectations of teacher efficacy and second language teacher and learner characteristics* (Unpublished master's thesis). Tarbiat Modarres University, Iran, Tehran.
- Koziol, S. M., & Burns, P. (1987). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *Journal of Educational Research*, 79(4), 205-209.
- Kuiper, R. & Pesut, D. (2004). Promoting cognitive and metacognitive reflective reasoning skills in nursing practice: Self-regulated learning theory. *Journal of Advanced Nursing*, 45, 381-391.
- Lanyon, L., & Hubball, H. (2008). Gender considerations and innovative learning-centered assessment practices. *Transformative Dialogues: Teaching and Learning Journal*, 2(1), 1-11.
- Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. National Comprehensive Center for Teacher Quality.
- Loreman, T., Sharma, U., & Forlin, C. (2013). Do pre-service teachers feel ready to teach in inclusive classrooms? A four country study of teaching self-efficacy. *Australian Journal of Teacher Education*, 38(1), 27-44.
- Mackey, A. & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, N.J.: Lawrence Erlbaum.
- Madsen, A. (2005). *Where is the "self" in teacher self-assessment? An examination of teachers' reflection and assessment practices in relation to their teaching practices*. Ph.D. dissertation. Iowa State University. Retrieved from <http://lib.dr.iastate.edu/rtd>
- Manouchehri, A. (2002). Developing teaching knowledge through peer discourse. *Teaching and Teacher Education*, 18, 715-737.

- McNamara, G., & O'Hara, J. (2008). The importance of the concept of self-evaluation in the changing landscape of education policy. *Studies in Educational Evaluation, 34*, 173-179.
- Nilsson, P. (2012). What do we know and where do we go? Formative assessment in developing student teachers' professional learning of teaching science, *Teachers and teaching: Theory and Practice, 19*(2), 37-41.
- Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (pp. 1-49). Greenwich, CT: JAI Press.
- Rahimi, S., Ayati, M., & Asgari, A. (2013). The relationship between teacher core self-evaluation, classroom management and educational achievement. *International Journal of Research in Organizational Behavior and Human Resource Management, 1*(4), 210-217.
- Rastegar, M. & Memarpour, S. (2009). The relationship between emotional intelligence and self-efficacy among Iranian EFL teachers. *System, 37*(4), 700-707.
- Regehr, G., & Eva, K. (2006). Self-assessment, self-direction, and the self-regulating professional. *Clinical Orthopaedics And Related Research, 449*, 34-38.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. Hedges (Eds.), *The handbook of research synthesis*. (pp. 231-244). New York: Russell Sage Foundation.
- Ross, A., & Starling, M. (2005, April). *Achievement and self-efficacy effects of self-evaluation training in a computer-supported learning environment*. Presented at the annual meeting of the American Educational Research Association, Montreal.
- Sabers, D., Cushing, K., & Berliner, D. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal, 20*(1), 63-88.
- Sargeant, J. (2008). Toward a common understanding of Self-assessment. *The Journal of Continuing Education in the Health Professions, 28*(1), 1-4.
- Sargeant, J., Armson, H., Chesluk, B., Dornan, T., Eva, K., Holmboe, E., Lockyer, J., Loney, E., Mann, K., van der Voluten, C. (2010). The processes and dimensions of informed self-assessment: A conceptual model. *Academic Medicine, 85*(7), 1212-1220.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

- Syafar, A. (2014). Self-assessment of certified EFL teachers in central Sulawesi on their professional competence. *Journal Pendidikan Humaniora Hal, 1*(3), 304-319.
- TLRP (2006, November). *Factors that make people more effective across their careers*. No. 20. Retrieved from TLRP.org
- Tschannen-Moran & Woolfolk-Hoy (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing the illusive construct. *Teaching and Teacher Education, 17*, 783-805.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2002, April). *The influence of resources and support on teachers' efficacy beliefs*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and teacher Education, 23*, 944-956.
- Westburg, J., & Jason, H. (2001). *Fostering reflection and providing feedback: Helping others learn from experience*. New York, NY: Springer.
- Woolfolk-Hoy, A., & Burke-Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*(4), 343-356.
- Worrell, F. & Kuterbach, L. (2001). The use of student ratings of teacher behaviors with academically talented high-school students. *Journal of Secondary Gifted Education, 12*(4), 236-247.
- Zarei, A. & Afshari, N. (2012). Experienced and novice Iranian teachers' perceptions as to the effect of intrinsic factors on teacher efficacy. *Basic Research Journal of Education Research and Review, 1*(1), 4-14.
- Zell, E. & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111-125.

Appendices

Appendix 1

Instructor Version of Teachers' Sense of Efficacy Scale (long form) Teacher Beliefs

Directions: This questionnaire is designed to help us gain a better understanding of Iranian EFL University instructors' teaching efficacy. Please indicate your opinion about each of the statements below. Your answers are confidential.

How much can you do?

Nothing (1) Very Little (2) Some influence (3) Quite A Bit (4) A Great (5)

1. How much can you do to get through to the most difficult students? (1) (2) (3) (4) (5)
2. How much can you do to help your students think critically?
(1) (2) (3) (4) (5)
3. How much can you do to control disruptive behavior in the classroom?(1) (2) (3) (4) (5)
4. How much can you do to motivate students who show low interest in school work?
(1) (2) (3) (4) (5)
5. To what extent can you make your expectations clear about student behavior?
(1) (2) (3) (4) (5)
6. How much can you do to get students to believe they can do well in school work?
(1) (2) (3) (4) (5)
7. How well can you respond to difficult questions from your students?
(1) (2) (3) (4) (5)
8. How well can you establish routines to keep activities running smoothly?
(1) (2) (3) (4) (5)
9. How much can you do to help your students value learning?
(1) (2) (3) (4) (5)
10. How much can you gauge student comprehension of what you have taught?
(1) (2) (3) (4) (5)
11. To what extent can you craft good questions for your students?
(1) (2) (3) (4) (5)
12. How much can you do to foster student creativity?
(1) (2) (3) (4) (5)
13. How much can you do to get students to follow classroom rules?
(1) (2) (3) (4) (5)
14. How much can you do to improve the understanding of a student who is failing?
(1) (2) (3) (4) (5)
15. How much can you do to calm a student who is disruptive or noisy?
(1) (2) (3) (4) (5)
16. How well can you establish a classroom management system with each group of students?
(1) (2) (3) (4) (5)
17. How much can you do to adjust your lessons to the proper level for individual students?
(1) (2) (3) (4) (5)
18. How much can you use a variety of assessment strategies?
(1) (2) (3) (4) (5)
19. How well can you keep a few problem students from ruining an entire lesson?
(1) (2) (3) (4) (5)
20. To what extent can you provide an alternative explanation/example when students are confused?(1) (2) (3) (4) (5)
21. How well can you respond to defiant students?
(1) (2) (3) (4) (5)
22. How much can you tap into the power of “selfhood”: encouraging students to pursue their own interests, develop their own perspectives, and express their values and dreams? (1) (2) (3) (4) (5)
23. How well can you implement alternative strategies in your classroom?
(1) (2) (3) (4) (5)
24. How well can you provide appropriate challenges for very capable students?
(1) (2) (3) (4) (5)

Experience in university instruction:years

Specialist field:**Appendix 2****Student Version of Teachers' Sense of Efficacy Scale (long form) Teacher Beliefs**

Directions: This questionnaire is designed to help us gain a better understanding of Iranian EFL University instructors' teaching efficacy. Please indicate your opinion about each of the statements below. Your answers are confidential.

How much can your instructor do?

Nothing (1) Very Little (2) Some influence (3) Quite A Bit (4) A Great (5)

1. How much can your instructor do to get through to the most difficult students?

(1) (2) (3) (4) (5)

2. How much can your instructor do to help his/her students think critically?

(1) (2) (3) (4) (5)

3. How much can your instructor do to control disruptive behavior in the classroom?

(1) (2) (3) (4) (5)

4. How much can your instructor do to motivate students who show low interest in school work?

(1) (2) (3) (4) (5)

5. To what extent can your instructor make his/her expectations clear about student behavior?

(1) (2) (3) (4) (5)

6. How much can your instructor do to get students to believe they can do well in school work?

(1) (2) (3) (4) (5)

7. How well can your instructor respond to difficult questions from his/her students?

(1) (2) (3) (4) (5)

8. How well can your instructor establish routines to keep activities running smoothly?

(1) (2) (3) (4) (5)

9. How much can your instructor do to help his/her students value learning? (1)

(2) (3) (4) (5)

10. How much can your instructor gauge student comprehension of what he/she have taught? (1)

(2) (3) (4) (5)

11. To what extent can your instructor craft good questions for his/her students? (1)

(2) (3) (4) (5)

12. How much can your instructor do to foster student creativity? (1)

(2) (3) (4) (5)

13. How much can your instructor do to get students to follow classroom rules? (1)

(2) (3) (4) (5)

14. How much can your instructor do to improve the understanding of a student who is failing? (1)

(2) (3) (4) (5)

15. How much can your instructor do to calm a student who is disruptive or noisy? (1)

(2) (3) (4) (5)

16. How well can your instructor establish a classroom management system with each group of students?

(1) (2) (3) (4) (5)

17. How much can your instructor do to adjust his/her lessons to the proper level for individual students?

(1) (2) (3) (4) (5)

18. How much can your instructor use a variety of assessment strategies? (1)

(2) (3) (4) (5)

- 19. How well can your instructor keep a few problem students from ruining an entire lesson?
(1) (2) (3) (4) (5)
- 20. To what extent can your instructor provide an alternative explanation/example when students are confused?
(1) (2) (3) (4) (5)
- 21. How well can your instructor respond to defiant students? (1)
(2) (3) (4) (5)
- 22. How much can your instructor tap into the power of “selfhood”: encouraging students to pursue their own interests, develop their own perspectives, and express their values and dreams?(1) (2) (3) (4) (5)
- 23. How well can your instructor implement alternative strategies in his/her classroom? (1)
(2) (3) (4) (5)
- 24. How well can your instructor provide appropriate challenges for very capable students? (1)
(2) (3) (4) (5)

Instructor’s name:

Appendix 3

- 1-It happens that an instructor’s assessment of his/her effectiveness is far away from his students’ ratings of their teaching. What could the reasons be?
- 2- Which aspect of your teaching do you assess more; Classroom management, instructional strategies, or student engagement? Why?
- 3- How has your teaching effectiveness been influenced by experience?
- 4-You might have had some mostly experienced instructors who talk about their memories instead of teaching at class, or do not correct exam papers; what do you think the reasons of their losing motivation might be?
- 5- What feedback do you get from university administrators, colleagues, and society regarding your work? How is your performance influenced by their feedback?
- 6- How do you avoid bias in your SA?
- 7-Do you think you have the ability to assess your instructional effectiveness?