Iranian English Language Teachers' Perceptions of Monitoring and Scaffolding Practices of Assessment for Learning: A Focus on Gender and Class Size

Mohammad Reza Nasr, Assistant Professor, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran
nasr_mr@yahoo.com
Mohammad Sadegh Bagheri, Assistant Professor*, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran
bagheries@gmail.com
Firooz Sadighi, Professor, Department of English Language, Shiraz Branch, Islamic Azad University, Shiraz, Iran
firoozsadighi@yahoo.com

Abstract

Recent innovations in formative assessment have turned the spotlight on the implementation of assessment for learning in the classroom. Notwithstanding a considerable wealth of research on assessment for learning in mainstream education, few research studies in the field of language teaching thus far have touched upon assessment for learning. This quantitative study investigated Iranian English language teachers' perceived monitoring and scaffolding practices in respect of their gender and class size. To achieve this purpose, 384 Iranian EFL teachers who were selected using convenience sampling completed a 28-item Likert scale questionnaire on assessment for learning entailing two main constructs, namely monitoring and scaffolding. Our findings revealed a statistically significant gender difference with regard to perceived scaffolding. Likewise, the results showed that EFL teachers' perceived monitoring and scaffolding practices did not differ with respect to class size. The key implications of the findings for the application of scaffolding and monitoring practices in the classroom were also addressed.

Keywords: English language teachers, gender, monitoring, scaffolding, class size, Assessment for learning

Introduction

A severe consequence of high-stakes exams is the possible adverse washback effect on teaching and learning (Berry, 2008). To Stiggins (2005), this assessment mode is not instrumental in enhancing learning at the classroom level. To fend off this issue, the notion of assessment for learning (henceforth, AFL) has been put forward (Wei, 2017) whose purpose is to use assessment during instruction to benefit learning (Broadfoot et al., 2002). This is suggestive of assessment bridge wherein assessment, teaching, learning, and curriculum are integrated to enhance learning (Colby-Kelly & Turner, 2007).In the wake of this integration, assessment for learning has gained in popularity in the context of language teaching (Harlen, 2006). An underlying premise of AFL is the awareness of the link between teaching and learning and more importantly, how information gleaned from assessments can inform teaching and learning (Lee, 2007).

As teachers' perceptions of AFL influence their actual assessment practices in classroom settings (Davison, 2004), it is of paramount importance to undertake a study on variables affecting their perceptions of AFL. Furthermore, high-quality AFL practices are almost rarely implemented in classroom settings and teachers might only employ some AFL practices non-systematically (Wiliam, 2011).

Despite extensive studies on AFL in mainstream education, there is a dearth of research on AFL in the context of language teaching (Rea-Dickins, 2004). Thus, more studies on AFL may lead to a better understanding of variables affecting its implementation and also an eager embrace of AFL within classrooms. Integral to AFL are two major dimensions, namely scaffolding and monitoring entailing classroom practices in the service of learning (Pat-El, Tillema, Segers, & Vedder, 2013). Adopting these practices within classrooms assists EFL teachers to orchestrate teaching, learning, and assessment.

Despite considerable research on the impact of class size on student achievement, there is little empirical data on the relationship between class size and teachers' assessment practices (Locastro, 2001; Reynolds, Reagin, & Reinshuttle, 2001). Therefore, it remains unclear whether teachers with smaller class sizes have expressed any preference for classroom assessment practices that would reflect an assessment for learning environment (Duncan & Noonan, 2007). Evidently, class size appears to be a common issue for teachers. The present study contributes to the literature by investigating variables including gender and class size that might influence the implementation of AFL within the classroom context.

Review of Literature
Assessment for Learning, Monitoring, and Scaffolding

AFL refers to assessments undertaken by teachers during instruction in order to identify student problems, make plans for upcoming steps in teaching, and give them feedback to enhance their learning process (Stiggins, 2005). Indeed, AFL supports learning in two ways. In the first place, teachers can adjust instruction based on evidence and make modifications that will lead to student learning. Second, students can use evidence of their current level of attainment to actively direct their learning (Chappuis & Chappuis, 2008).

During AFL, learners take an essential role during their learning (Wang, 2010) as they work collectively with their teacher and classmates to communicate their evidence of learning. As a result, they are actively engaged in self/ peer-assessment (Alexander, 2013). AFL involves an interactive and learner-oriented assessment approach (Chen, May, Klenowski, & Kettle, 2014). To Popham (2008), AFL is seen as the integration of assessment and instruction as a continuous process which assists teachers in modifying and adjusting their teaching. Students can use assessment information to modify their learning as well. To Stiggins (2008), during assessment for learning, teachers give students ongoing formative feedback. That is, AFL arises from an assessment culture with scaffolding at its heart that guides instruction to promote learning and makes students independent learners (Black & Wiliam, 1998). The underlying tenet of AFL is that assessment can enhance student learning (Stiggins, 2002a) occurring in real-life contexts (Darling-Hammond, 2010). In point of fact, AFL is the interaction between the teacher and the student where students perform active roles in learning processes (Stiggins, 2002b). During assessment for learning, students learn most effectively when they become aware of the success criteria (Stiggins, 2006). In essence, AFL is not an approach to assessment but a process conducive to active learning and teaching (Clarke, Hattie, & Timperley, 2003). It is an approach to classroom-based assessment serving the purpose of providing opportunities for developing self-regulated learners (Black & Wiliam, 1998). The central objective of AFL is to engage learners in self-assessment in a way that they recognize where they stand in the learning cycle, realize where they are required to be next, and how to get there (James et al., 2007). An underlying principle of assessment for learning is that students decide on important matters in the classroom (Berry, 2008).This highlights the central role of learner autonomy in an AFL approach (Tsagari, 2016).What it all boils down to is that AFL is a teaching style in the form of an

informal assessment embedded in the course of instruction (Black, Harrison, Lee, Marshall, & William, 2003).

Scaffolding as a way of operationalizing Vygotsky's concept of ZPD refers to various kinds of support that learners receive from teachers to foster learning. In this study, by scaffolding is meant teachers' practices empowering learners to recognize areas for improvement (Stiggins, 2005). Scaffolding during assessment provides teachers with valuable insights to better diagnose learners' potentials in view of future instruction that is closely associated with the notion of dynamic assessment (Booth, 2012). According to Pat-EL et al. (2013), scaffolding embraces teacher-driven instruction centering on practices related to learning objectives, assessment criteria, and questioning strategies. Effective scaffolding entails strategies that allow language learners to get involved in challenging and instructional activities. However, it does not necessarily encompass the simplification of the language or the content (Quiocho & Ulanoff, 2009). Scaffolding contributes to formative assessment meaning that teachers assess student's strengths and weaknesses and guide their learning by providing hints rather than answers to give them the chance to get to the answer on their own (Aydeniz, 2009). To Allal and Ducrey (2000), formative assessment intends to scaffold students in their ZPD. In Brookhart's view (2003) during formative assessment, students are given scaffolded assistance to optimize their performance.

Monitoring progress is not an end but a complement to lesson planning performed to assess the effectiveness of classroom activities, strategies, and teaching materials aiming to identify learners' strengths and weaknesses (Cotton, 1998). Monitoring is a teaching practice adopted by teachers to observe students' progress by checking for understanding, providing feedback, and aligning teaching with student learning (Frey & Fisher, 2011). Monitoring practices including planning and the evaluation of learning processes are subsumed under metacognitive strategies (Berry, 2008). These strategies are in essence central to assessment as learning, a concept closely related to AFL (Clark, 2012). Some educators hold that assessment for learning embraces the ideas related to assessment as learning (Earl & Katz, 2006). However, AFL foregrounds teacher-mediated learning while assessment as learning is premised on student-led learning (Clark, 2012). In an AFL setting, teachers keep track of their learners through observation and assessment of their learning over time. Teachers try to communicate with their students in various manners, for instance, by motivating them to think about the ways they can enhance the language learning process, and by reflecting on their progress in learning (Oz, 2014). Fundamental to monitoring are practices dealing with learners' strong and weak points in learning (Lee & Mak, 2014). In a similar vein, Pat-El, Tillema, Segers, and Vedder (2015) maintain that monitoring is concerned with keeping track of student learning progress to stimulate their self-monitoring to fight off challenges to improve learning.

Studies on AFL in the Context of English Language Teaching

A few studies have addressed AFL in relation to teachers' demographics and other related variables in the context of foreign or second language teaching (e.g., Hasan & Zubairi, 2016, Oz, 2014; Nasr, Bagheri, Sadighi, & Rassaei, 2018; Nasr, Bagheri, Sadighi, & Rassaei, 2019). One of the first studies of this kind was performed by Oz (2014), who investigated EFL teachers' perceptions of assessment for learning. The participants of his study were 120 Turkish EFL teachers selected from public and private institutions i.e. high schools and universities. The study also examined teachers' perceived monitoring and scaffolding practices regarding such variables as years of service, gender, and the context of teaching. The results of his study suggested that Turkish EFL teachers' monitoring practices differed significantly in relation to years of teaching experience, gender, and the context of teaching.

In a study by Hasan and Zubairi (2016), Malaysian ESL teachers' AFL practices were identified. The sample of their study was 120 teachers who were asked to respond to an AFL instrument adapted from Pat-EL et al. (2013). They examined differences in monitoring and scaffolding practices regarding gender. Based on their results, ESL teachers extensively adopted monitoring and scaffolding practices in classroom settings. As for the effect of gender, it was found that no statistically significant differences existed in terms of scaffolding and monitoring practices.

Nasr et al. (2018) explored Iranian EFL teachers' perceived scaffolding and monitoring practices. They also investigated EFL teachers' perceived monitoring and scaffolding with respect to their years of service, educational degree, and English language proficiency levels taught. Their results did not suggest any significant differences concerning the demographic characteristics in question.

Nasr et al. (2019) studied EFL teachers' perceived monitoring and scaffolding practices concerning textbooks taught and teaching context. They also delved into barriers to the implementation of AFL. Conducting a mixed methods research study, they revealed that EFL teachers' perceived monitoring and scaffolding practices differed as regards the textbooks taught and the context of teaching i.e. language institutes and high schools.

To investigate the impact of class size on teachers' classroom assessment preferences, Gonzales and Aliponga (2012) conducted a study on 116 English teachers including Japanese English language teachers working (n=61) in the Philippines and English teachers teaching English in Japan (n=55). They compared their assessment preferences and concluded that teachers' classroom assessment preferences varied significantly with different class sizes. They also found that teachers with smaller class sizes achieved lower scores while teachers with big class sizes scored higher.

The four studies reviewed above explored EFL teachers' perceived scaffolding and monitoring practices of AFL and examined the practices in question with respect to variables associated with teachers e.g. years of teaching experience, academic degrees, and language proficiency levels taught, teaching context and textbooks taught. As evident, AFL has not been investigated with respect to class size. One advantage of our study is its rather large sample size. The studies performed on AFL in relation to gender (e.g., Hasan & Zubairi, 2016, Oz, 2014, did not take language institute EFL teachers into consideration. These concerns highlight the need to investigate AFL relating to the variables in question. To fill the lacuna, we studied these variables in terms of the two AFL constructs i.e. scaffolding and monitoring. This study will, therefore, address the following research questions:

Q1. Is there a significant difference between Iranian EFL teachers' perceived monitoring and scaffolding practices with respect to their gender?

Q2. Is there a significant difference between Iranian EFL teachers' perceived monitoring and scaffolding practices with regard to class size?

Methodology

A quantitative research design was adopted to address the two research questions of the study. Data obtained from a self-report AFL questionnaire were analyzed quantitively. According to Creswell (2015), the survey research study is utilized when the researcher intends to analyze perceptions, behaviors, and population characteristics in the form of variables within a study.

Participants and Setting

A total of 384 Iranian EFL teachers from language institutes and high schools selected based on convenience sampling were involved in the current study. Given the unknown population of EFL teachers in Iran, 384 EFL teachers were determined according to Krejcie and Morgan's Table (1970) with the 95% of the confidence level and the degree of accuracy of 0.05. The participants' ages ranged from 25 to 60 years. EFL teachers involved in the study were from two educational settings i.e. language institutes and high schools across Iran. Teachers' profiles are presented in Table 1.

Table1. *EFL Teachers' Profiles*

Charateristics	Description	N	%
Gender	Male	185	48.2
	Female	199	51.8
Educational setting	Institutes	203	52.9
	High schools	181	47.1
Academic degree	BA	135	35.1
	MA	169	44.0
	Ph.D./Ph.D. candidate	80	20.8
Years of teaching experience	1–5 years	82	21.3
	6–10 years	80	20.8
	11–15 years	68	17.7
	16–20 years	77	20.0
	+20 years	77	20.0
Class size	Large	97	25.2
	Average	141	36.7
	Small	146	38.0

Data Collection Procedures

The data were gathered through an AFL questionnaire administered on a social media network called Telegram where the required data could be captured in a short period of time from a large sample of EFL teachers across Iran. To achieve this, a questionnaire invitation was forwarded to Telegram groups whose members were EFL teachers. They expressed their willingness to take part in the study by clicking the link to complete the questionnaire taking them 5 minutes. The data were gathered over a period of 10 months.

Instrument

A validated questionnaire (see Appendix A) was adapted from Pat-El et al. (2013). The first part of the questionnaire elicits teachers' demographics as well as their class size. Regarding the class size, the participants were required to select from three categories of small (up to 15 language learners), average (16 to 25 language learners), and large (more than 25 language learners). The second part entails 28 statements gauging two central dimensions of AFL, namely monitoring and scaffolding constituting 12 and 16 items, respectively on a 5-point Likert scale. It is noteworthy to mention that very few changes were made to the questionnaire to suit language teaching context i.e. the addition of the words" English" and "language institutes" to the questionnaire where appropriate. The questionnaire items subjected to changes were then

reviewed by two experts in the field of applied linguistics and their comments were taken into account to establish the face validity of the questionnaire.

Piloting the Instrument

To ensure reliability, we pilot-tested the questionnaire with 50 English language teachers. The internal consistency was calculated using Cronbach's alpha for the entire instrument (α = .857) and monitoring (α = .826), and scaffolding (α = .778). Given the reliability coefficients obtained, the instrument was deemed reliable to be administered to EFL teachers.

Data Analysis

Descriptive and inferential statistics were performed for the data analysis using IBM SPSS Statistics 24.

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was run via AMOS Graphics 24 in the form of structural equation modeling to evaluate the model fit of assessment for learning questionnaire as a measurement model.

In this study, six major goodness-of-fit indices as depicted in Table 2 were used to check the model fit of the AFL questionnaire. The fit indices values of the model demonstrate a moderately good fit of the questionnaire to the data verifying the factor structure of the questionnaire.

Table 2. *Fit Indices of the Measurement Model*

	Fit indices						
	RMSEA	CFI	TLI	IFI	GFI	df	X^2/df CMIN/DF
Default model	.033	.941	.936	.941	.915	349	1.413
Saturated model	-	1.000	-	1.000	1.000	0	-
Acceptable range	<.08	>.90	>.90	>.90	>.90	-	< 3

Results

Results for the First Research Question

Is there a significant difference between Iranian EFL teachers' perceived monitoring and scaffolding practices as a function of their gender?

The first objective of the study was to determine the possible significant difference between Iranian EFL teachers' perceived monitoring and scaffolding in terms of their gender. To probe the effect of gender on teachers' perceived monitoring and scaffolding practices, we conducted an independent samples *t*-test. The pertaining results are depicted in Tables 3 and 4.

Table 3. *Descriptive Statistics of Males' and Females' Perceived Monitoring and Scaffolding Practices*

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Monitoring	Male	186	4.2886	.35528	.02605
	Female	198	4.3516	.34927	.02482

Scaffolding	Male	186	4.3275	.37449	.02746
	Female	198	4.4032	.34655	.02463

As shown in Table 3, the mean score of females' perceived scaffolding was 4.4 while the males' perceived scaffolding was calculated to be 4.32. Also, according to the results of the descriptive statistics, the mean scores of females' perceived monitoring and males' perceived monitoring appeared to be 4.35 and 4.28, respectively.

Table 4. *Independent Samples t-test to Compare Males' and Females' Perceived Monitoring and Scaffolding*

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Monitoring	Equal variances assumed	.021	.884	1.752	382	.081	-.06300	.03596	.13371	.00771
	Equal variances not assumed			1.751	379.588	.081	-.06300	.03598	.13375	.00775
Scaffolding	Equal variances assumed	.755	.386	2.057	382	.040	-.07569	.03680	.14804	-00334
	Equal variances not assumed			2.052	374.670	.041	-.07569	.03689	.14822	-00316

The levels of statistical significance presented in Table 4, revealed that males and females were significantly different concerning perceived scaffolding (sig. = .04, $p < .05$). The eta squared statistic (.01) reflects a small effect size. Based on the mean scores presented in Table 3, females (Mean= 4.40) held a more positive perception of scaffolding than males (Mean= 4.32). The results of the independent samples *t*-test also revealed that males and females were not significantly different concerning their perceived monitoring (sig. = .08).

Results for the Second Research Question

Is there a significant difference between Iranian EFL teachers' perceived monitoring and scaffolding practices with regard to the class size?

To answer the second research question, a one-way MANOVA was performed on the two dependent variables of the study i.e. monitoring and scaffolding. As explained previously, the independent variable of class size consisted of three groups of small, average, and large size.

The assumptions of one-way MANOVA were examined to ensure it is the right statistical analysis to address the second research question. To this end, multivariate normality was performed by conducting Mahalanobis distance using regression analysis. The value of Mahalanobis distance turned out to be 2.057 which is lower than the critical value of 13.82 for studies with two dependent variables indicating that the multivariate normality requirement of MANOVA was met.

Multicollinearity assumption was checked by performing correlations to check the strength of the correlation between monitoring and scaffolding. Correlations up around .8 and .9 are reasons for concern. The correlation appeared low (.355) satisfying the assumption of multicollinearity.

The homogeneity of variance-covariance matrices was also examined using Box's M Test of Covariance Matrices. P values for Box's Test should be larger than .001 (Tabachnick & Fidell, 2007). The sig. value obtained (.784) as displayed in Table 5 was larger than .001 suggesting that the observed covariance matrices of the two dependent variables of the study were equal across groups.

Table 5. *Box's Test of Equality of Covariance Matrices*

Box's M	3.217
F	.532
df1	6
df2	1745555.286
Sig.	.784

Next, the Levene's test (Table 6) was run, the results of which were not significant (p > .05) in terms of scaffolding and monitoring variable indicating that equal variances were assumed.

Table 6. *Levene's Test of Equality of Error Variances*

	F	df1	df2	Sig.
Scaffolding	.844	2	381	.431
Monitoring	1.311	2	381	.271

The descriptive statistics of EFL teachers from different class sizes are presented in Table 7.

Table7. *Descriptive Statistics of EFL Teachers' Scaffolding and Monitoring concerning Class Size*

Class size		Mean	Std. Deviation	N
Scaffolding	Large	4.3557	.37894	97
	Average	4.3487	.37022	141
	Small	4.3910	.34276	146
	Total	4.3665	.36187	384

Monitoring	Large	4.3061	.37091	97
	Average	4.3107	.36019	141
	Small	4.3412	.33526	146
	Total	4.3211	.35314	384

The results of multivariate tests as depicted in Tables 8 indicate that no statistically significant differences were found between EFL teachers' perceptions of scaffolding and monitoring in terms of class size f (4,76) =.352, Wilks' Lambda = .843.

Table 8. *Multivariate Tests*

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.995	4.024E4[a]	2.000	380.000	.000	.995
	Wilks' Lambda	.005	4.024E4[a]	2.000	380.000	.000	.995
	Hotelling's Trace	211.776	4.024E4[a]	2.000	380.000	.000	.995
	Roy's Largest Root	211.776	4.024E4[a]	2.000	380.000	.000	.995
Class size	Pillai's Trace	.004	.352	4.000	762.000	.843	.002
	Wilks' Lambda	.996	.351[a]	4.000	760.000	.843	.002
	Hotelling's Trace	.004	.351	4.000	758.000	.844	.002
	Roy's Largest Root	.004	.683[b]	2.000	381.000	.506	.004

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

Table 9. *Tests of Between-Subjects Effects*

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Scaffolding	.144[a]	2	.072	.547	.579	.003
	Monitoring	.096[b]	2	.048	.384	.682	.002
Intercept	Scaffolding	7071.449	1	7071.449	53874.213	.000	.993
	Monitoring	6923.851	1	6923.851	55340.380	.000	.993
Class size	Scaffolding	.144	2	.072	.547	.579	.003

	Monitoring	.096	2	.048	.384	.682	.002
Error	Scaffolding	50.009	381	.131			
	Monitoring	47.668	381	.125			

a. R Squared = .003 (Adjusted R Squared = -.002)

b. R Squared = .002 (Adjusted R Squared = -.003)

In addition, the MANOVA tests of between-subjects effects performed on all dependent variables separately (Table 9) revealed that EFL teachers' perceived monitoring and scaffolding were not statistically significant in terms of class size. However, based on Table 7, the mean scores of EFL teachers' perceived monitoring (Mean=4.39) and scaffolding (Mean=4.34) in small size classes turned out to be marginally higher than those of the average and large size classes.

Discussion

The purpose of this study was to examine Iranian EFL teachers' perceptions of assessment for learning practices with respect to their gender and class size.

The findings of the first research question revealed a significant difference between female and male EFL teachers as regards their perceived scaffolding. However, the findings demonstrated a marginal gender difference in monitoring practices in favor of female EFL teachers. Overall, many studies have included gender as a potential variable in classroom assessment, but it is rarely reported as showing any effect. However, this study suggests that there was a marked tendency for female EFL teachers toward the implementation of AFL practices particularly scaffolding practices. This may be attributed to the nature of females showing a great deal of support, understanding, and sensitivity towards their students. They are more nurturing and responsive to their students than are their male colleagues. Further, they are more likely to employ cooperative strategies, use metacognitive skills as well as multiple strategies to help them practice tasks. Such strategies and practices, however, are integral to scaffolding and monitoring practices of assessment for learning. In this regard, in a study investigating the reading comprehension scaffolding, Salem (2017) concluded that female EFL teachers tended to be more diligent in enabling their learners to perform each component of a task and to assure that they have mastered it

The findings of the current study in terms of gender run counter to the study conducted by Oz (2014) in that no significant difference was reported as a function of gender. Likewise, he concluded that male EFL teachers obtained a higher mean score in monitoring practices and lower mean score in scaffolding practices. However, the participants in his study included 120 EFL teachers working in primary/ middle / high schools as well as universities. It seems that the AFL practices implemented by university instructors are not entirely identical to EFL teachers teaching English in high schools or language institutes. This is due to the fact that language learning curriculum is different in the contexts in question. On the other hand, in a study by Estaji and Fassihi (2016), no statistically significant relationship was found between EFL teachers' use of formative assessment practices and gender. Investigating teachers' preferences for assessment, Han and Kaya (2014) found that there was no significant difference between male and female teachers' assessment preferences including assessment as learning, summative assessment, AFL, assessment for instruction, and assessment to inform.

The results of the second research question demonstrated that EFL teachers' perceived monitoring and scaffolding did not significantly differ as a function of class size. The indication

is that Iranian EFL teachers embark on assessment for learning practices regardless of the number of language learners in the classroom. Another line of explanation might be that AFL practices are infused into teachers' regular instruction. This finding is in harmony with that of Duncan and Noonan (2007) who found that class size did not play a central role in teachers' assessment preferences. Our results concerning class size are in conflict with the findings of Noori, Shafie, Mashwani, and Tareen (2017) in which large classes were viewed as a major factor impacting the use of formative assessment. Along the same lines, Biggs (1998) holds that large class sizes affect Asian teachers to implement formative assessment in the classroom. However, Danielson (2008) maintains that teachers tend to embed assessment practices into their teaching, implementing AFL, more specifically in classes with a fewer number of students. Gonzales and Aliponga (2012) noted that class size may impact EFL teachers' assessment preferences. They found that EFL teachers with a fewer number of students or smaller class sizes obtained lower scores than those teaching in big class sizes in terms of assessment preferences. Although our results did not show any statistical significance, the mean scores of teachers' monitoring and scaffolding turned out to be higher for small size classes than large size classes. One reason is that EFL teachers in small size classes have more time to employ AFL practices, in particular, providing language learners with one-to-one descriptive feedback.

Conclusions

The present study contributes to a greater understanding of AFL with respect to variables including gender and class size. This research study indicated that female EFL teachers revealed a markedly higher level of perceived scaffolding practices. Male EFL teachers are therefore recommended to adopt a severe orientation toward scaffolding to reap more benefits of its practices. Building on this, in-service programs on AFL can offer opportunities for male EFL teachers to integrate AFL practices into their instruction.

Another finding of this study was that EFL teachers perceived monitoring and scaffolding practices did not show substantial differences in terms of class size, yet revealing insignificantly marginal instances of both monitoring and scaffolding practices in classes with a fewer number of language learners. Thus, it is suggested to set limits on the number of language learners in English classes. The small size classes can help EFL teachers divert more attention to language learners by providing them with descriptive feedback and by creating more room for an interactive classroom environment.

The current study holds significant implications for pre/ in-service teacher education programs, teacher educators, and researchers. The results can conduce to researchers to explore new approaches to implementing AFL as a group in settings where the option of reducing class sizes is not applicable.

The implications of this study could be the change in language teaching methodologies in schools and language institutes and the development of new syllabuses which would integrate AFL into teachers' instruction of language skills in the future.

The findings of this study carry significant implications for school administrators and language institute managers. The study also contributes to the implementation of AFL by raising EFL teachers' awareness of assessment for learning.

The main limitation of the current study lies in the self-report method of collecting data. The participants of this study may have not given their honest responses to the questionnaire items. This requires a high level of participant motivation to complete the questionnaire (Mitchell & Jolley, 2010). Hence, the findings of the current study should be interpreted with caution.

Despite the limitation discussed above, as one of the first attempts centered on AFL in Iran, the study provides insights into EFL teachers' perceptions of assessment for learning in relation to gender and class size. Yet, qualitative studies i.e. classroom observations are called for to shed more light on male and female teachers' AFL behaviors in different class sizes.

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1-3.

Allal, L., & Ducrey, G. P. (2000). Assessment of—or in—the zone of proximal development. *Learning and Instruction*, *10*(2), 137-152.

Aydeniz, M. (2009). Formative assessment: Improving learning in secondary classrooms. *School Science and Mathematics*, *109*(7), 428-431.

Berry, R. (2008). *Assessment for learning*. Aberdeen: Hong Kong University Press.

Biggs, J. (1998) The assessment scene in Hong Kong. In P. Stimpson & P. Morris (Eds) *Curriculum and assessment for Hong Kong: two components, one system* (pp.214-315). Hong Kong: Open University of Hong Kong Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for learning-putting it into practice*. Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Booth, D. (2012). Scaffolding during the formal assessment of young EAL learners. *Australian Review of Applied Linguistics*, *35*(1), 5-27.

Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge: University of Cambridge. School of Education.

Chappuis, S., & Chappuis, J. (2008). The best value in formative assessment. *Educational Leadership, 65*(4), 14-19.

Chen, Q., May, L., Klenowski, V., & Kettle, M. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy & Practice*, *21*(3), 271-285.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205-249.

Clarke, S., Hattie, J., & Timperley, H. (2003). *Unlocking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom*. New Zealand Edition. Wellington: Printlink Ltd.

Colby-Kelly, C., & Turner, C. E. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review, 64*(1), 9-37.

Cotton, K. (1988). *Monitoring student learning in the classroom*. Portland. OR: Northwest Regional Educational Laboratory.

Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson

Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning*. Washington, DC: Council of Chief State School Officers.

Danielson, C. (2007. Assessment for learning: For teachers as well as students. In C.A. Dwyer (Ed.). *The Future of Assessment: Shaping Teaching and Learning* (pp. 191-213). New York: Taylor & Francis.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, *21*(3), 305-334.

Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, *53*(1).

Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind.* Winnipeg, Manitoba: Western Northern Canadian Protocol.

Estaji, M., & Fassihi, S. (2016). On the relationship between the implementation of formative assessment strategies and Iranian EFL teachers' self-efficacy: Do gender and experience make a difference? *Journal of English Language Teaching and Learning, 8*(18), 65-86.

Frey, N., & Fisher, D. (2011). *The formative assessment action plan: Practical steps to more successful teaching and learning.* Alexandria, VA: Association for Supervision & Curriculum Development.

Gonzales, R., & Aliponga, T. (2012). Classroom assessment preferences of Japanese language teachers in the Philippines and English language teachers in Japan. *MEXTESOL Journal*, 36(1).

Han, T., & Kaya, H. I. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education, 4*(1), 77-93.

Harlen, W. (2006). The role of teachers in the assessment of learning. *Pamphlet produced by the Assessment Systems for the Future projects (ASF) Assessment Reform group, UK*, 347-364.

Hasan, W. M. M. W., & Zubairi, A. M. (2016). Assessment competency among primary English language teachers in Malaysia.Prague: *Proceedings of MAC-ETL 2016*, 66.

James, M., McCormick, R., Black, P., Carmichael, P., Drummond, M. J., Fox, A., & Swaffield, S. (2007). *Improving learning how to learn: Classrooms, schools, and networks.* Routledge.

Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *Canadian Modern Language Review, 64*(1), 199-213.

Lee, I., & Mak, P. (2014). Assessment as learning in the language classroom. *Assessment as learning. Hong Kong: Education Bureau.*

Locastro, V. (2001). Teaching English to large classes. TESOL Quarterly, 35(1), 493-496.

Mitchell, M. L., & Jolley, J. M. (2010). *Research design explained.* Belmont, CA: Wadsworth Cengage Learning.

Nasr, M., Bagheri, M. S., Sadighi, F., & Rassaei, E. (2018). Iranian EFL teachers' perceptions of assessment for learning regarding monitoring and scaffolding practices as a function of their demographics. *Cogent Education*, *5*(1), 1558916.

Nasr, M., Bagheri, M. S., sadighi, F., & Rassaei, E. (2019). Iranian EFL teachers' assessment for learning Practices and barriers: Do textbooks taught and teaching context matter? *Cogent Arts & Humanities*, *6*(1), 1646691.

Noori, A., Shafie, N. H., Mashwani, H. U., & Tareen, H. (2017). Afghan EFL lecturers' assessment practices in the classroom. *Imperial Journal of Interdisciplinary Research, 3*(10), 130–143.

Oz, H. (2014). Turkish Teachers' practices of assessment for learning in English as a foreign language classroom. *Journal of Language Teaching & Research, 5*(4).

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology, 83*(1), 98-113.

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2015). Multilevel predictors of differing perceptions of Assessment for Learning practices between teachers and students. *Assessment in Education: Principles, Policy & Practice, 22*(2), 282-298.

Popham, W. (2003). *Test better, teach better: The instructional role of assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Quiocho, A. L., & Ulanoff, S. H. (2009). *Differentiated literacy instruction for English language learners.* Boston, MA: Allyn & Bacon.

Rea-Dickins, P. (2004*). Understanding teachers as agents of assessment. Language Testing, 21(3), 249–258.*

Reynolds, A., Reagin, M., & Reinshuttle, K. (2001). Less is more: What teachers say about decreasing class size and increasing learning. *American School Board Journal, 199*(9), 30-32.

Salem, A. A. M. S. (2017). Scaffolding reading comprehension skills. *English Language Teaching, 10*(1), 97-111.

Stiggins, R. J. (2002a). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758-765.

Stiggins, R. J. (2002b). Assessment for learning. *Education Week, 21*(26), 30-33.

Stiggins, R. J. (2006). Assessment for learning: A key to student motivation and learning. *Phi Delta Kappa EDGE*, 2(2), 1-19

Stiggins, R. J. (2008). Correcting errors of measurement that sabotage student learning. *The future of assessment: Shaping teaching and learning, New York, NY: Erlbaum.*

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Wang, T. H. (2010). Web-based dynamic assessment: Taking assessment as teaching and learning strategy for improving students' e-Learning effectiveness. *Computers & Education*, *54*(4), 1157-1166.

Wei, W. (2017). A critical review of washback studies: Hypothesis and evidence. In Al-Mahrooqi, R., Coombe, C., Al-Maamari, F., & Thakur, V. (Eds.), *Revisiting EFL Assessment: Critical Perspective* (pp. 49-67*).*Cham*:* Springer.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14.

Appendix A

Dear Instructor:

You are invited to complete a 28- item questionnaire aiming to explore English language teachers' perceptions of assessment for learning practices. The questionnaire is ANONYMOUS and no personally identifiable information will be collected. Your participation in this research is entirely VOLUNTARY.

Kindest Regards

Part A: Demographic information.

1. Gender: Female   ☐ Male

2. Academic degree earned:

☐ Bachelors   ☐ Masters   ☐ Doctoral candidate/PhD   ☐ other (please specify)

3. Please select ONE educational institution at which you teach English the most.
☐ English language institute   ☐ High school

4. The class size (the number of language learners in a class) which you teach the most.
☐ Small (1-15)   ☐ Average(16-25)   ☐ Large(More than 25)

Part B: **EFL Teachers' Perceptions of assessment for learning practices**
Instructions:
　　Using the rating scale provided below, please indicate how much you agree or disagree with each statement. Choose the response that comes closest to describing your opinion concerning scaffolding and monitoring practices of assessment for learning.

1= Strongly disagree (SD)　2= Disagree (D)　3= Not sure (NS)　4= Agree (A) 5= Strongly agree (SA)						
I	**Perceived monitoring**	SD 1	D 2	NS 3	A 4	SA 5
1	I encourage my students to reflect upon how they can improve their language learning assignments.					
2	After a test, I discuss the answers given with each student.					
3	While working on their language learning assignments, I ask my students how they think they are doing.					
4	I involve my students in thinking about how they want to learn English at a language institute/high school.					
5	I give my students the opportunity to decide on their language learning objectives.					
6	I ask my students to indicate what went well and what went badly concerning their assignments.					
7	I encourage students to reflect upon their learning processes and how to improve their learning.					
8	I inform my students about their strong points concerning language learning.					
9	I inform my students about their weak points concerning language learning.					
10	I encourage my students to improve their language learning processes.					
11	I give students guidance and assistance in their language learning.					
12	I discuss language learning tasks with my students to help them understand the content better.					
13	I discuss with my students the progress they have made in learning English.					
14	After an assessment, I inform my students on how to improve their weak points.					
15	I discuss with my students how to utilize their strengths to					

		SD 1	D 2	NS 3	A 4	SA 5
	improve on their language learning tasks.					
16	Together with my students, I consider ways on how to improve on their weak points.					
II	**Perceived scaffolding**	SD 1	D 2	NS 3	A 4	SA 5
17	I adjust my language teaching whenever I notice that my students do not understand a topic.					
18	I provide my students with guidance to help them gain an understanding of the content taught.					
19	During my class, students are given the opportunity to show what they have learned.					
20	I ask questions in a way my students understand.					
21	I am open to student contribution in my language class.					
22	By asking questions during class, I help my students gain an understanding of the content taught.					
23	I allow my students to ask each other questions using English during class.					
24	I ensure that my students know what areas they need to work on in order to improve their results.					
25	I give my students opportunities to ask questions.					
26	My students know what the evaluation criteria for their work are.					
27	I ensure that my students know what they can learn from their assignments.					
28	I can recognize when my students reach their language learning goals.					