

Turkish Teachers' Practices of *Assessment for Learning* in the English as a Foreign Language Classroom

Hüseyin Öz

Department of Foreign Language Education, Faculty of Education, Hacettepe University, Ankara, 06800, Turkey

Abstract—Recently there has been a noticeable paradigm shift in educational assessment where assessment and student learning are viewed as inseparable and assessment is perceived as a tool for supporting student learning. This study was designed to investigate Turkish teachers' preferences of common assessment methods in the English as a foreign language (EFL) classroom, their Assessment for Learning (AFL) practices, and determine whether they differed in their AFL practices according to some variables such as years of teaching experience, gender, and public vs. private school context. 120 EFL teachers completed the online self-report Assessment for Learning Questionnaire for Teachers (TAFL-Q) consisting of 28 statements on a 5-point (ranging from strongly disagree to strongly agree) Likert scale ($\alpha = .92$). The findings revealed that most Turkish EFL teachers rely on conventional methods of assessment rather than formative assessment processes. While they generally reported high levels of perceived monitoring (82.86%) and scaffolding practices (86.94%) of AFL, they had significant differences in their assessments, especially monitoring in support of student learning, according to years of teaching experience, gender, and private vs. public schools variables ($p < .05$). To practice AFL, teachers need to appraise their current assessment practices and develop AFL strategies and feedback procedures. Their perceptions of AFL cannot change overnight, nor do they can do it individually and independently. They need support from different sources to recognize the effect of their previous perspectives on their practices and weigh them against the insights offered by the new assessment culture. Through self-report and observational data from both teachers and students to get better insights into monitoring and scaffolding practices, further research could be conducted to explore probable mismatches between teacher and student perceptions of AFL in EFL classrooms.

Index Terms—assessment for learning, monitoring, scaffolding, English language teaching, English as a foreign language (EFL) teachers

I. BACKGROUND

In recent years there has been a noticeable change in educational assessment where assessment and student learning are viewed as inseparable and assessment is perceived as a tool for supporting student learning (Davison and Leung, 2009; Stiggins, 2008). This change in assessment paradigm is due to some reform actions in various settings around the world to improve educational planning and practices (Assessment Reform Group, 1999, 2002; Black and Wiliam, 1998, Chow and Leung, 2011; OECD, 2005; Phakiti and Roever 2011). As this policy-supported practice is increasingly encouraged globally, English language teachers are also called on to apply suitable assessment procedures to monitor and evaluate student progress in their own classrooms. (Davison & Leung, 2009).

Given that assessment plays a crucial role in raising standards of achievement as well as informing and supporting student learning, *assessment for learning* (AFL) has been introduced in contrast to *assessment of learning* (AOL) in classrooms and schools (Assessment Reform Group, 2002; Black and Wiliam, 1998; Black et al., 2003; Stiggins, 2005, 2008). To put it another way, there is now a shift from a more traditional testing culture where instruction and assessment are perceived as independent from each other, towards an assessment culture where assessments are conducted, both formally and informally, through monitoring and scaffolding, aligned with classroom instruction to enhance student learning (Black & Wiliam, 1998, Black et al., 2003; Davison & Leung, 2009).

AFL is referred to as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” (Assessment Reform Group, 2002, p. 2). Thus, it is an interactive and learner-centered assessment approach evolving in many subject areas of education including English language teaching (Chow & Leung, 2011; Chen et al., 2013). In contrast to AOL, utilized for reporting and decision-making purposes and thus making summative judgments of learning outcomes, AFL integrates assessment into instruction as an ongoing process, where teachers use assessment information to make adjustments in their instructional endeavors and resources. Learners then use this information to modify their learning processes accordingly (Popham, 2008; Pat-El, Tillema and Segers, 2013). To sum up, based on the constructivist views of learning (Berry, 2008), AFL aims to better understand how learners learn, what they can do or cannot do, and makes some deliberations and decisions on how to help them learn.

In field of English language teaching, there is no widely accepted definition of this assessment approach associated with constructivism, with numerous terms used almost interchangeably to mean similar practices and procedures, including terms such as *formative assessment*, *teacher-based assessment*, *classroom-based assessment*, *school-based assessment*, *dynamic assessment*, and *alternative assessment*. These emphasize different aspects of the assessment process, suggesting “a more teacher-mediated, context-based, classroom-embedded assessment practice, explicitly or implicitly defined in opposition to traditional externally set and assessed large scale formal examinations used primarily for selection and/or accountability purposes” (Davison & Leung, 2009, p. 395). Although AFL is grounded in the principles of formative assessment, AFL puts emphasis on everyday “progress in learning as students climb the curricular scaffolding leading up to state standards (Stiggins, 2005, p. 328). Thus, for the purposes of this paper AFL is used to mean much more than formative assessment, highlighting the shift in assessment culture that takes into account monitoring and scaffolding learners, aligning language instruction in support of student learning, and providing students with greater autonomy in the learning process (Black & Wiliam, 1998; Pat-El et al., 2013; see Berry, 2008 for further details about assessment *of, for, and as* learning).

The current literature on assessment shows that different researchers highlight various AFL principles (e.g., Assessment Reform Group, 2002; Black & Wiliam, 1998). This study focuses on two issues of AFL: monitoring learners to follow their progress and scaffolding to enable learners to understand in what areas they should improve (Stiggins, 2005; Pat-El et al., 2013). Numerous studies and reports (Assessment Reform Group, 2002; OECD, 2005; Davison & Leung, 2009; Chow & Leung, 2011; Darling-Hammond, 2010) explain and emphasize these two dimensions in promoting student learning. Teachers who implement AFL can monitor their students by observing and making assessments of what is happening in the classroom during learning activities. Then, they can engage in various interactions with students, for example, by encouraging them to reflect on how they can improve their language learning, discussing with students the progress they have made in learning English, and informing students on their weak and/or strong points. As far as scaffolding is concerned, teachers and students need to follow a strategy where they engage in appropriate social interactions, with teachers providing “demonstrations, support, guidance and input and gradually withdrawing these as the learner becomes increasingly independent” (Richards and Schmidt, 2010, p. 507).

There is evidence emerging in the international literature that teachers’ conceptions and perceptions of assessment affect their instruction and assessment practices (Black & Wiliam, 1998; Davison & Leung, 2009; Brown, 2008). To find out the impact of assessment on student learning, some studies have focused on conceptual attitudes about assessment practices and perceived/observed practice (Brown, 2008; Pat-El et al., 2013). Brown (2008) showed that teachers’ conceptual attitudes towards assessment influence their assessment practices, and in turn, determine learning outcomes. Few studies have attempted to investigate teachers’ perceptions of assessment that affect their instruction and assessment practice. The findings of a recent study (Pat-El et al., 2013) revealed that perceptions of assessment situations and interactions held by teachers and students should be harmonious in order to improve the effects of assessment on student learning. In a very recent study, Lee and Coniam (2013) investigated the implementation of AFL for EFL writing within an examination-driven AOL system in Hong Kong. Their findings indicated that despite efforts to plan and pre-assess instruction, share learning goals with students and use feedback forms to provide feedback and align assessment with instruction, teachers had to follow conventional practices that required detailed attention to errors and summative scores, and were unable to engage students in multiple drafting and peer evaluation on a regular basis.

Research reports and studies into language teachers’ perceptions and practices of classroom-based assessment is very limited (Chow & Leung, 2011; Wu, 2013; Zhou, 2013) and there is a call for research into this type of assessment and assessment research (Davison & Leung, 2009). In the context of Turkey there has also been a shift in the conceptual framework guiding the EFL curriculum and instruction practices, resulting in a new curriculum document putting more emphasis on formative assessment. (MoNE, 2013). To the best of the researcher’s knowledge, whether EFL teachers apply appropriate assessment procedures in support of student learning remains unanswered. The current literature review revealed that there is not even a single study dealing with Turkish teachers’ perceptions and practices of AFL in the language classroom. Thus, seeing this gap in the literature, the present research grew out of a desire to contribute in this respect by investigating Turkish teachers’ perceived monitoring and scaffolding practices of AFL in the EFL classroom and find out factors that may explain probable differences among teachers. Thus, the following research questions were formulated to guide this study.

1. What methods of assessment do Turkish EFL teachers use in their classrooms?
2. To what extent do they monitor and scaffold their students to support learning?
3. Are there significant differences of assessment practices among them by
 - (a) years of teaching experience,
 - (b) private vs. public sector,
 - (c) school type,
 - (d) pre-service vs. in-service training, and
 - (e) faculty-college graduated from?
4. Do they differ in their perceived monitoring and scaffolding practices of AFL in the EFL classroom in relation to their gender and sector?

Given the time and resource limitations, these questions were addressed by using quantitative analysis of data drawn from practitioners teaching EFL in both public and private educational institutions. In this study, it is expected that exploring teachers' perceptions and practices of AFL in the EFL classroom will provide valuable data for both in-service and pre-service training as well as administrators involved. By providing insights into teachers' perceptual attitudes, the findings should lead teacher trainers to evaluate teachers' practices and make adjustments in training programs in line with the curriculum guidelines of their institutions.

II. METHOD

A. Participants

A purposeful sampling technique was used to collect responses from as many online teachers as possible. Patton (1990) describes this technique as the process of choosing specific "information-rich cases" from which the researcher can learn significant information related to the research. Thus, criterion sampling was employed to choose participants based on certain characteristics, specifically, EFL teachers currently teaching in public and private educational institutions. To obtain the most representative sample possible, the questionnaire was sent to as many online EFL teachers as possible, using a social networking service. This way, it was possible to reach and connect with target audience more effectively. The online survey was administered in May 2013 to 400 online teachers. A pre-notification message was first sent out informing potential participants of the survey, followed by a second message containing a link to the instrument. Two follow-up reminders were then sent out to respondents over the course of a month. A total of 120 responses were gathered, yielding a 30% response rate for the survey. This rate was considered satisfactory and higher than most web-based surveys (Manfreda, Bosnjak, Berzelak, Haas & Vehovar, 2008; Shih & Fan, 2008). Accordingly, the sample included 120 EFL teachers working in public and private educational institutions. 80% of the participants (N=96) graduated from EFL teacher education departments that offered a language 'testing' course in their curriculum. 20% of the participants (N=23) graduated from language departments at faculties of science and letters, including English Language and Literature, American Culture and History, and English Linguistics. Except graduates of EFL teacher education departments, the participants reported that they had received an EFL certificate from a Turkish university. As indicated in Table 1, most EFL teachers had 1 to 5 years of teaching experience, followed by those with 6-10 years of teaching experience.

TABLE 1
BACKGROUND INFORMATION ABOUT THE PARTICIPANTS

Institution/School Type	Frequency	Percentage	Years of Teaching Experience	Frequency	Percentage
Primary School (Grades 1-4)	15	12.5	1-5	62	51.70
Middle School (Grades 5-8)	30	25	6-10	37	30.8
High School	24	20	11-15	10	8.3
University	51	42.5	16+	11	9.2
Total	120	100	Total	120	100
Education Sector	Frequency	Percentage	Gender	Frequency	Percent
Private	39	32.5	Female	87	72.5
Public	81	67.5	Male	33	27.5
Total	120	100	Total	120	100
Pre-service Training in Language Assessment	Frequency	Percentage	In-service Training in Language Assessment	Frequency	Percent
Yes	73	60.8	Yes	65	54.2
No	47	39.2	No	55	45.8
Total	120	100	Total	120	100

B. Instrument

The instrument for this study was a self-report questionnaire consisting of two parts. The first part included questions that characterized the participants such as gender, years of teaching experience, the types of educational institution and sector where they taught EFL, and common assessment methods they used, whereas the second part included Assessment for Learning Questionnaire for Teachers (TAFL-Q) adapted from Pat-El et al. (2013). They developed and validated the instrument in a very recent study that included a sample of 237 teachers from all disciplines, ranging from arts to sciences. It consisted of 28 statements divided into two subscales: (1) perceived monitoring (16 items) and (2) perceived scaffolding (12 items). The participants rated the statements on a 5-point (ranging from strongly disagree to strongly agree) Likert scale. It did not contain any "negative" or "unfavourable" statements that could otherwise be reverse-coded in statistical analysis. Internal consistency was examined using Cronbach's alpha for the whole instrument ($\alpha = .92$) as well as the two subscales, perceived monitoring ($\alpha = .88$) and perceived scaffolding ($\alpha = .89$). These high alpha values indicated good internal consistency and satisfactory psychometric properties of the items in the TAFL-Q (Field, 2009).

C. Data Collection and Analysis

The data for the study were gathered using an online self-report questionnaire administered through a social networking service. The potential respondents were sent out information about the purpose of the study and requested to

participate by clicking the link to the instrument. The online implementation ensured that there were no cases with unreliable or missing data responses among all collected data. Data analysis was performed to address the research questions formulated previously. Statistical data analyses were conducted using IBM SPSS Statistics 21, a comprehensive computer program used to help researchers perform statistical analysis quickly and accurately. Descriptive statistics such as frequency and mean were computed to characterize the collected data. The independent samples T-test was run to compare the two level variables such as gender and sector. The independent-samples t-test is employed when the two groups or sets of scores whose means are being compared are independent of each other (Mackey and Gass, 2005; Field, 2009). The two way multivariate analysis of Anova (two-way MANOVA) was conducted to measure the main and interaction effects of independent variables on the EFL teachers' perceived monitoring and scaffolding practices of assessment within the AFL framework. In fact, MANOVA is used to test the hypothesis that one or more independent variables, or factors, exercise an effect on a set of two or more dependent variables. This is done to avoid the probability of the so-called Type I error in data analysis. In addition, wherever MANOVA results demonstrated significant differences between groups, Tuckey's Post hoc test was run to determine the exact location of differences among groups measured in the study.

III. RESULTS

The present study yielded some significant findings regarding Turkish teachers' AFL practices in EFL classrooms. This section presents the results of the study in terms of descriptive and inferential statistics, structured along with the research questions set in the study.

A. Research Question 1: What Methods of Assessment Do Turkish EFL Teachers Use in Their Classrooms?

To obtain a model for presenting the results of teachers' preferred assessment methods, their preferences were divided into three groups according to the frequency of use and their percentages. The assessment methods were considered as 'the most preferred' ones if their percentages were within 71 to 100, 'the second most preferred' if they were within 41 to 70, and 'the least preferred' ones if they were within 0 to 40 of the total range.

The results of descriptive analyses revealed that fill in the blank, multiple-choice, true-false, matching and short answer exams were among the most preferred type of assessment methods employed by EFL teachers. Overall, 46.18% of them prefer to use these methods in assessing their language learners. The results of percentages analyses showed that nearly nine in ten (89.17%) of teachers are more likely to use fill in the blank method, 86.67% prefer to use multiple choice type of assessment method, 82.50% favor true-false test items, 81.67% prefer matching method, and 79.17% prefer to use short-answer exam type of assessment method (Figure 1).

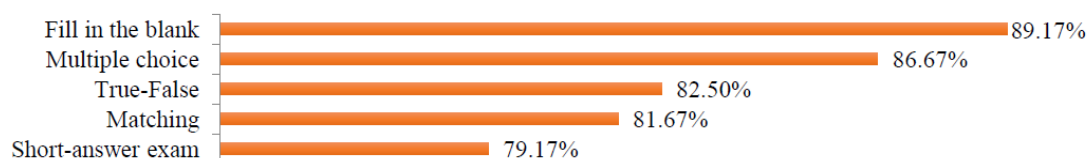


Figure 1. The Most Preferred Assessment Methods by EFL Teachers

Oral exams, group work, project, portfolio, performance assessment, essay type exam and oral presentation were the second most preferred assessment methods among EFL teachers. Their frequency of use and percentages ranged from 60.83% for oral exam to 46.67% for oral presentation. Other assessment methods in this category fell within these two methods with nearly the same frequency and percentages of usage, i.e. around 50% (Figure 2). Only 40 teachers are more likely to use formative ways of assessment.

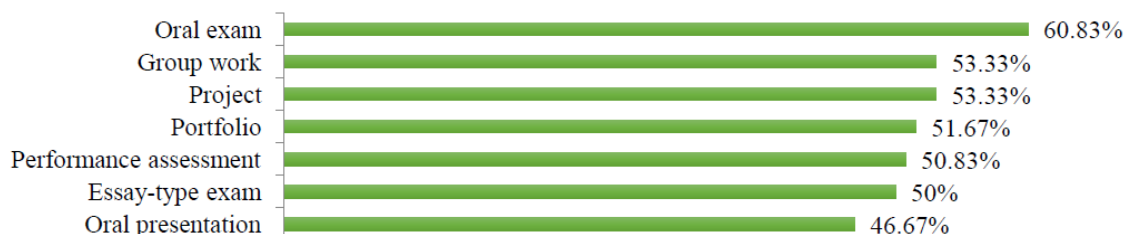


Figure 2. The Second Most Preferred Assessment Methods by EFL Teachers

The least preferred assessment methods were rubric, self- and peer-assessment, observation form, drama, structure grid and other methods which, on average, account for 14.27% of the total percentages of methods used for assessing language learners (Figure 3). These methods are mostly considered as formative practices (Black et al. 2003, Chen et al., 2013).

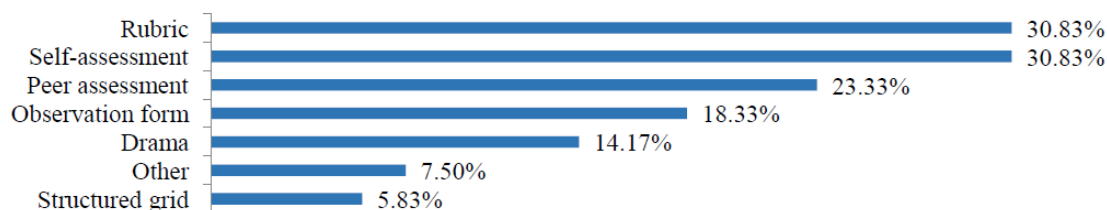


Figure 3. The Least Preferred Assessment Methods by EFL Teachers

B. Research Question 2: To What Extent Do They Monitor and Scaffold Their Students to Support Learning?

The results of descriptive analyses indicated that the mean scores obtained for the items under perceived monitoring factor were consistent within the scale. That is, 81.25% of the participants rated above 4 (Table 2). The highest mean score ($M=4.58$, $SD=.71$) was received for item 11 (*I give students guidance and assistance in their language learning*), suggesting that most teachers (94.16%) prefer to monitor students during their practical teaching. In contrast, the lowest mean score was ascribed to item 5 (*I give my students the opportunity to decide on their language learning objectives*), indicating less participation on the part of students in the process of language learning as well as assessment. The careful scrutiny of the results indicate that all the mean scores under 4, i.e. scores for items 3, 4 and 5, are related to learners' active involvement in the process of language learning as well as assessment. This implies that EFL teachers are less inclined to involve their students in decision making and setting language learning objectives. On the whole, 82.86% of them agreed with perceived monitoring aspect of AFL, 3.7% disagreed and 13.54% were undecided on the usefulness of monitoring as an AFL practice.

TABLE 2
DESCRIPTIVE STATISTICS FOR EFL TEACHERS' PERCEIVED MONITORING PRACTICES OF AFL

Items and Item Descriptions	1	2	3	4	5	Mean	SD
I. Perceived Monitoring							
1. I encourage my students to reflect upon how they can improve their language learning.	1	1	13	46	59	4.34	.77
2. After a test, I discuss the answers given with each student.	1	7	19	36	57	4.18	.95
3. While working on their assignments, I ask my students how they think they are doing.	0	5	25	60	30	3.96	.79
4. I involve my students in thinking about how they want to learn English at school.	0	4	26	58	32	3.98	.78
5. I give my students the opportunity to decide on their language learning objectives.	3	13	40	39	25	3.58	1.01
6. I ask my students to indicate what went well and what went badly concerning their assignments.	2	4	22	54	38	4.02	.88
7. I encourage students to reflect upon their learning processes and how to improve their learning.	0	3	13	56	48	4.24	.74
8. I inform my students on their strong points concerning language learning.	1	4	12	50	53	4.25	.83
9. I inform my students on their weak points concerning language learning.	0	6	13	50	51	4.22	.83
10. I encourage my students to improve on their language learning processes.	1	0	7	53	59	4.41	.68
11. I give students guidance and assistance in their language learning.	1	2	4	33	80	4.58	.71
12. I discuss assignments with my students to help them understand the content better.	1	1	9	43	66	4.43	.74
13. I discuss with my students the progress they have made in learning English.	1	2	10	55	52	4.29	.76
14. After an assessment, I inform my students on how to improve their weak points.	2	1	15	42	60	4.31	.84
15. I discuss with my students how to utilize their strengths to improve on their assignment.	1	2	17	47	53	4.24	.82
16. Together with my students, I consider ways on how to improve on their weak points.	0	1	15	53	51	4.28	.71

Note: 1 = Strongly disagree, 5 = Strongly agree.

The results of descriptive analyses indicated that the mean scores for the items under perceived scaffolding of AFL were also highly consistent within the scale. Strangely enough, all of the participants rated the items positively and the obtained mean scores were all above 4 (Table 3). All in all, 86.94% of them agreed to adopt scaffolding, 3.19% disagreed and 9.87% were hesitant whether to use it or other way round. The highest mean score ($M=4.65$, $SD=.65$) was received for item 25 (*I give my students opportunities to ask questions*). This suggests that most (96%) believe in the importance of asking questions during learning a foreign language. The lowest mean score, on the other hand, was obtained for item 24 (*I ensure that my students know what areas they need to work on in order to improve their results*). This implies that not all teachers pay much attention to the problematic areas during their teaching, with 79% zeroing in on the weak points of students.

TABLE 3
DESCRIPTIVE STATISTICS FOR EFL TEACHERS' PERCEIVED SCAFFOLDING PRACTICES OF AFL

Items and Item Descriptions	1	2	3	4	5	Mean	SD
II. Perceived Scaffolding							
17. I adjust my language teaching whenever I notice that my students do not understand a topic.	1	3	5	48	63	4.41	.76
18. I provide my students with guidance to help them gain understanding of the content taught.	1	3	12	45	59	4.32	.82
19. During my class, students are given the opportunity to show what they have learned.	1	3	17	42	57	4.26	.85
20. I ask questions in a way my students understand.	1	1	4	31	83	4.62	.67
21. By asking questions during class, I help my students gain understanding of the content taught.	1	0	10	41	68	4.46	.72
22. I am open to student contribution in my class.	1	1	5	33	80	4.58	.69
23. I allow my students to ask each other questions using English during class.	2	5	10	42	61	4.29	.91
24. I ensure that my students know what areas they need to work on in order to improve their results.	0	7	18	56	39	4.06	.84
25. I give my students opportunities to ask questions.	1	1	3	29	86	4.65	.65
26. My students know what the evaluation criteria for their work are.	1	3	18	50	48	4.18	.83
27. I ensure that my students know what they can learn from their assignments.	2	4	23	45	46	4.08	.92
28. I can recognize when my students reach their language learning goals.	0	3	17	62	38	4.13	.74

Note: 1 = Strongly disagree, 5 = Strongly agree.

C. Research Question 3: Are There Significant Differences of Assessment Practices among Them by (a) Years of Teaching Experience, (b) Private vs. Public Sector, (c) School Type, (d) Pre-service vs. In-service Training, and (e) Faculty-college Graduated From?

The two-way multivariate analysis of variance (two-way MANOVA) was conducted to assess the effects of such factors as years of teaching experience, private vs. public sector, school type, pre-service vs. in-service training, and faculty-college graduated from on the participants' tendency to choose and adopt monitoring and scaffolding practices of AFL in EFL classrooms. A non-significant Box's M test ($p=0.823$, $P>0.05$) indicated that there was homogeneity of covariance matrices of the dependent variables across the levels of groups.

The results of multivariate tests indicated a significant main effect for participants' years of teaching experience with respect to their assessment practices (Wilks' Lambda=.648, $F(6, 66) = 2.664$, $P=0.022$, $p<.05$, partial $\eta^2 = 0.195$, observed power=.830). However, the results showed no significant main effect between gender, school, sector, pre-service vs. in-service training and faculty-college factors and preference for specific assessment practices among EFL teachers (Table 4). There was also statistically significant interaction effect between gender and sector variables on the teachers' tendency to use their assessment practices (Wilks' Lambda=.730, $F(2, 33) = 6.109$, $P=0.006$, $p<.05$, partial $\eta^2 = 0.270$, observed power=.857).

TABLE 4
THE EFFECT OF INDEPENDENT VARIABLES MEASURED ON THE CHOICE OF ASSESSMENT METHODS

Variables	Main Effect				
	Λ	df	F	Sig.	η^2
Gender	.997	2;33	.045	.956	.090
Years of teaching experience	.648	4;66	2.664	.022*	.195
School	.907	6;66	.548	.770	.047
Sector	.966	2;33	.574	.569	.034
Pre-service training	.937	2;33	1.103	.344	.063
In-service training	.979	2;33	.358	.702	.021
Faculty- college	.855	2;33	2.142	.133	.115

* Significant at .05 level

The results of tests of between-subjects effects, or univariate tests, demonstrated that there was a significant difference among EFL teachers in relation to years of teaching experience and adopting monitoring activities of AFL $F(3,85) = 4.141$, $p=0.013$; $p < .05$, partial $\eta^2 = .27$, whereas no significant difference was found as regards their scaffolding practices. Also, there was a significant interaction effect between gender and sector variables and preference for monitoring $F(3, 85) = 12.076$, $p=0.001$; $p < .05$, partial $\eta^2 = .26$. However, gender and sector variables had no significant interaction effect on participants' perceived scaffolding in support of student learning (Table 5).

TABLE 5
RELATIONSHIP BETWEEN INDEPENDENT VARIABLES AND ASSESSMENT METHODS

Variables Independent	Dependent	Univariate Analysis			
		df	F	Sig.	η^2
Years of teaching experience	Monitoring	3,85	4.141	.013 [*]	.27
	Scaffolding	3,85	.486	.694	.041
Gender *Sector	Monitoring	3,85	12.076	.001 [*]	.26
	Scaffolding	3,85	3.315	.077	.089

* Significant at .05 level

A post-hoc test was also run to find out exactly where the significant differences between groups existed. The post hoc (multiple) comparisons using the Tuckey's HSD test revealed that the participants with 1-5 years of teaching experience differed significantly in monitoring students during language learning from those with 6-10 years of teaching experience (Table 6).

TABLE 6
POST HOC TEST FOR TEACHERS' PREFERENCES OF ASSESSMENT METHOD

Tuckey's HSD	Multiple Comparisons				
Variables	(I) Years of teaching experience	(J) Years of teaching experience	Mean Difference (I-J)	Std. Error	Sig.
Monitoring	1-5 years	6-10 years	4.3204 [*]	1.36985	.017
		11-15 years	.8258	2.24710	.983
		16+ years	.1804	2.15735	1.000

* Significant at .05 level

Analysis of the reliability effect test, according to Cohen (1988), showed a moderate significant difference for years of teaching experience (partial $\eta^2 = .27$) and interaction effect of gender and sector variables (partial $\eta^2 = .26$) on EFL teachers' perceived monitoring and scaffolding practices of AFL in supporting student learning. Cohen's (1988) reliability effect indexes for small, moderate, and strong relationships are $r=0.1$, 0.30 , and 0.50 , respectively. As for the main and interaction effect of other factors on assessment preferences, the results showed no significant differences between the variables.

D. Research Question 4: Do They Differ in Their Perceived Monitoring and Scaffolding Practices of AFL in the EFL Classroom in Relation to Their Gender and Sector?

The results of descriptive statistics for Turkish EFL teachers' perceptions and practices of assessment for learning (AFL) in the EFL classroom demonstrated differences between males and females with respect to their preferences to use monitoring and/or scaffolding for assessing their students. The highest mean score ($M= 67.51$; $SD=8.20$) for perceived monitoring was received by males while the lowest mean score ($M=67.22$; $SD=7.58$) was observed in female group. Conversely, females rated more positively ($M=52.19$; $SD=6.63$) than males ($M=51.54$; $SD= 5.65$) in their perceptions of perceived scaffolding factor. However, the results of the Independent Samples t-test showed no significant differences between groups in relation to their perceived monitoring, $t(118)=-.176$, $P=0.861$, $P>0.05$, and perceived scaffolding, $t(118)=-.498$, $P=0.619$, $P>0.05$.

The results further revealed that 82.76% of the participants agreed with *Perceived Monitoring* practice of assessment, whereas four in ten (3.7%) reported not practicing it. Moreover, nearly nine in ten (86.94%) of them agreed with scaffolding, while 3.19% stated that they preferred not to adopt scaffolding as an assessment practice in classrooms. Drawing upon these findings, it can be concluded that Turkish EFL teachers are more likely to use scaffolding as an AFL practice. As shown in Figure 4, the total mean score for perceived monitoring method was 4.20 while the total mean score for perceived scaffolding was 4.33.

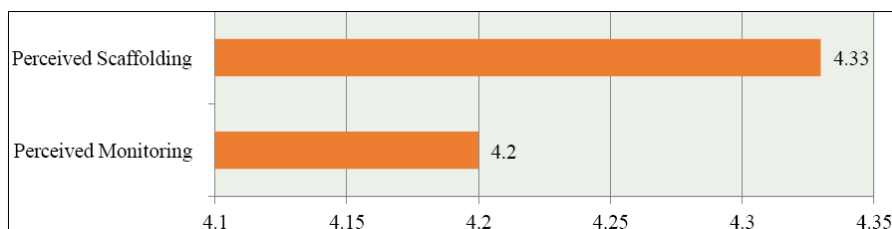


Figure 4. Turkish EFL Teachers' Perceived Practices of AFL in the EFL Classroom

As for the sector variable, the results, as determined by Independent samples t-test, showed a significant difference between teachers who work for private sector, $t(118)=2.253$, $P=0.026$, $P<0.05$ and those from public sector in relation to monitoring factor, while no significant difference was found between groups regarding scaffolding factor, $t(118)=1.426$, $P=0.157$, $P>0.05$. The highest mean scores for both monitoring ($M = 69.61$, $SD=7.54$) and scaffolding ($M=53.20$, $SD= 5.53$) was obtained for private sector, while the low scores for both monitoring ($M=66.19$, $SD=7.89$) and scaffolding ($M=51.44$, $SD=6.68$) were received for public schools.

IV. DISCUSSION

An important finding of the present research is that most Turkish EFL teachers rely on conventional methods of assessment rather than formative assessment tools such as self-assessment and peer assessment which are considered as the indispensable part of monitoring aspect of language teaching and learning. This may be due to the fact that EFL teachers consider formal examinations as to be the only form of assessment. In other words, formal or summative assessment, otherwise known as assessment of learning (AOL), is mainly practiced by EFL teachers rather than formative assessment or assessment for learning (AFL) (Boraie, 2012) due to the washback effect of final assessment or testing on language teaching, learning and assessment in contexts where students' performance in final exams is of utmost importance rather than their achievements as in the case of process learning and portfolio, project, and self- or peer-assessment. However, as Sardareh and Saad (2012, p. 344) rightly put it, in AFL learners play an active role in language learning and monitoring their progress. "They constantly collaborate with their teacher to monitor their current level of achievement in relation to the learning intentions." In other words, learners are given opportunities to actively communicate their learning evidence to their teachers, peers, and their parents, monitor their learning progress and control their success coupled with a strong self-reliance and self-confidence resulting in their success in language learning.

Focusing only on AOL clearly runs counter to recent developments in the assessment culture supported by experts as well as national and international organizations (OECD, 2005; Black et al., 2003; Pat-El et al., 2013). Monitoring student learning is also suggested by *Common European Framework of Reference (CEFR) for Languages: Learning, Teaching, Assessment* (CoE, 2001). Following CEFR, emphasis on alternative assessment methods *such as self-assessment, project and portfolio evaluation, (student-teacher cooperation), pen and paper tests, self-and peer evaluation* (including listening and speaking skills) and *teacher observation and evaluation* are also emphasized in the new English language curriculum put into effect by the Turkish Ministry of Education in 2013 for grades 2 to 8 (MoNE, 2013). Self-assessment is greatly emphasized by the new curriculum since learners/users are encouraged to monitor their own progress and achievement in the development of communicative competence as the ultimate goal of language learning.

Additionally, MoNE's (2013) new curriculum emphasizes that students' achievements are going to be controlled and evaluated through self-assessment checklists which ask students to assess their own learning from an action-based perspective. This does not mean, however, that formal or summative assessment should be totally abandoned when assessing learning outcomes. In keeping with the new curriculum, formal evaluation of the language learners will be carried out through the application of written and oral exams, quizzes, homework assignments and projects. This will be done, as in the past, to provide an objective record of students' success at the end of the semester and academic year. Therefore, both AFL and AOL are going to be put into practice in Turkish schools.

The significance of self- and peer-assessment as useful and widely-used strategies in AFL or formative assessment has also been highlighted by researchers interested in Assessment for Learning in language classrooms. Chow and Leung (2011, p. 142), for instance, argue that "one of the characteristics of AFL is that through the process of self- and peer-assessment students are expected to study the assessment criteria repeatedly and refine their judgment with the support of teacher feedback". They further state that the use of checklists in the project is that it provides ample opportunities for both teachers and students to use the same checklists for teacher, and student self- and peer-assessments, reducing the difference in expectations between students and teachers.

In an attempt to investigate the implementation of AFL for EFL writing within an examination-driven AOL system in Hong Kong, Lee and Coniam (2013) found that while teachers strengthened planning and pre-assessment instruction, involving students in making decisions on learning objectives, providing feedback and aligning assessment with instruction, they had to use conventional practices that required detailed attention to errors and summative scores, and failed to engage students in multiple drafting and peer evaluation on a regular basis. They also found that students had mixed opinions about motivation and this indicated that they were experiencing some tension between innovative and traditional assessment practices. Their results also confirmed that AFL contributed to the improvement of students' writing performance, suggesting that AFL had a facilitative role in language learning.

Another finding of this study revealed that Turkish EFL teachers differ significantly in their perceptions and practices of monitoring central to AFL to support student learning. Indeed, teachers with 1-5 years of experience differed in perceived monitoring practices from those with 6-10 years of experience. This is closely related to the development of teacher efficacy beliefs (Akbari and Tavassoli, 2011) among prospective teachers since 'there is some evidence that course work and practica have differential impacts on personal and general teaching efficacy (Hoy, 2000, p. 6). According to Hoy (2000), general teaching efficacy goes through stages of fluctuation and appears to increase during college coursework, then decline during student teaching. This suggests that the "optimism of young teachers may be somewhat tarnished when confronted with the realities and complexities of the teaching task (Hoy, 2000, P.6)". Nevertheless, additional observational data is required to confirm this aspect of Turkish EFL teachers' practices of monitoring as part of AFL. Also, it would be of interest to conduct interviews with students and collect data about their perceptions of teachers' assessment actions.

Although gender had no significant main effect on the participants' perceived monitoring and scaffolding practices, there was indeed a significant interaction effect between gender and sector on their perception of monitoring. A close

examination of the mean scores for the groups revealed that female teachers got lower scores in the monitoring indicator, which also requires additional data to confirm their monitoring practices in AFL. With regard to sector variable, the findings revealed that the participants working for private schools, are more likely to monitor and scaffold their students than teachers from public schools. This implies that there exist a myriad of factors ranging from availability of facilities, social support, teacher assessment literacy, teacher autonomy, and school accountability (Ajila, 1997; Brown, 2008) to fringe benefits- various non-wage compensations provided to employees in addition to their normal salaries. As a matter of fact, the purpose of offering fringe benefits mainly is to provide incentive for the employees, increase the economic security of them, award employees who are doing notably well or having seniority, give first chance for an individual for job promotion, attract employees to fields where there is a high demand such as teaching, keep quality of employees, etc (Stanford, 2008; Malveaux, 2010). It is also argued that rewards, whether intrinsic or extrinsic, do have important implications for individuals' overall job satisfaction and desire to commit to their job (Papadaki and Papadaki, 2006) and is one of the factors which help to combat burnout among employees.

While male EFL teachers reported lower scores as regards scaffolding practices, there were no significant differences of perceived scaffolding practices between male and female participants. In fact, both male and female teachers had high perceptions of scaffolding for AFL in support of student learning, suggesting that from a sociocultural perspective they value scaffolding as a crucial assessment practice in promoting student learning in Turkish context. This finding seems to be in line with those of another study by Kayi-Aydar (2013, p.324) that revealed that "student and teacher questions scaffolded language learning and use, and positively affected students' participation during teacher-led whole class interactions. Similarly, Sardareh and Saad (2012, p.351) argue that AFL, based on the teaching strategies like scaffolding and cooperative learning, "improves students' learning and guides them to develop their learning. Therefore, teacher should provide students with an appropriate learning context that help them improve their learning."

V. CONCLUSIONS

Notwithstanding the widespread adoption of AFL in education in the western context (Black et al., 2003; OECD, 2005), there has been comparatively little research into AFL in English as a second or a foreign language. That is, AFL has been neglected by teachers and researchers, especially in Turkish context, partly because of the traditional dominance of the field by great tendency to employ summative assessments to make decisions about students' learning and to measure the extent of their achievement of the instructional program learning outcomes (Boraie, 2012; Popham, 2009, Stiggins, 2005, 2008). Currently, the field of language assessment is undergoing a shift from conventional testing culture (AOL) to assessment culture (AFL), whereby assessment practices aim to inform not only the teachers about student learning but also students about their own achievement.

The present study showed that Turkish EFL teachers also differ significantly in their perceptions and practices of assessment for learning (AFL) in the EFL classroom, especially in employing monitoring their language learners' performance during the course. Although some actively use monitoring as a reliable assessment strategy for supporting student learning, there are a considerable number of them who pass it by indifferently for no good reasons. It is important for teachers to remember that AFL is different from traditional forms of examination and testing in that traditional forms consider context as a neutral and unimportant factor and suggests that the assessor or examiner must remain objective throughout the whole assessment process. However, the core content of AFL is that assessment should take place during regular curriculum between teachers and students since the ultimate attainment in language learning can and should be shaped and improved by teachers' and students' active involvement, and through monitoring, self- and peer- assessment, and constructive feedback on the part of the students. It is also worth mentioning that there is ample evidence in the literature (Davison & Leung, 2009, Carless, 2008; Davison & Hamp-Lyons, 2010; Kennedy et al., 2006), suggesting that even AOL for students' language skills should be carried out language formatively since this helps teachers to give constructive student feedback and enhance language learning.

On the other hand, one must accept that it is hardly possible for teachers to accomplish AFL by working individually or independently. There is a need to recognize their existing assessment practices and then "plan a comprehensive program that takes into account the interrelationships between teaching, learning, and assessment" (Lee, 2007, p. 209). Teachers can then determine instructional strategies and feedback procedures by considering the school's language curriculum. Language teachers must also ask their learners to take responsibility for their own learning by engaging in assessments, for example exercising self-assessment grid (CeO, 2001) on a regular basis. As AFL is an important professional skill for the 21st century teachers, both pre-service and in-service teacher trainers need to guarantee teachers' understanding and exercising it (Boraie, 2013; Davison & Leung, 2012; Lee, 2007).

A major limitation of this study is that the findings are based on self-report data only from the EFL teachers, who may have reported more use of AFL strategies than they actually did. It would be of great interest to gather self-report and observational data concurrently from both teachers and students to get better insights into monitoring and scaffolding practices and find out whether there are great teacher-student mismatches for AFL in EFL classrooms.

REFERENCES

- [1] Ajila, C. O. (1997). Job motivation and attitude to work as correlates of productivity among workers in manufacturing companies in Lagos State, Nigeria. (Unpublished PhD Dissertation). O.A.U Ile-Ife Osun State, Nigeria.
- [2] Akbari, K., & Tavassoli, K. (2011). Teacher efficacy, burnout, teaching style, and emotional intelligence: possible relationships and differences. *Iranian Journal of Applied Linguistics (IJAL)*, 14(2), 31-61.
- [3] Assessment Reform Group. (1999). Assessment for learning: Beyond the black box. Cambridge: University of Cambridge School of Education. Retrieved on 2 September 2013 from <http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-Beyond-the-Black-Box.pdf>
- [4] Assessment Reform Group. (2002). Assessment for learning: 10 principles. Retrieved on 2 September 2013 from <http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- [5] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 21(1), 49-97. doi:10.1080/0969595980050102
- [6] Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). Assessment for learning: Putting it into practice. Buckingham: Open University Press.
- [7] Brown, G. T. L. (2008). Conceptions of assessment: Understanding what assessment means to teachers and students. New York, NY: Nova Science Publishers.
- [8] Boraie, D. (2012). Formative vs. summative assessment: Does it matter? TESOL Connections, September 2013. Retrieved on 4 October 2013 from <http://newsmanager.commpartners.com/tesolc/issues/2012-09-01/3.html>.
- [9] Carless, D. (2008) Developing productive synergies between formative and summative assessment processes. In M. F. Hui & D. Grossman (Eds.), *Improving teacher education through action research* (pp. 9-23). New York: Routledge.
- [10] Chen, Q., May, L., Klenowski, V., & Kettle, M. (2013). The enactment of formative assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy & Practice (In press)* doi:10.1080/0969594X.2013.790308
- [11] Chow, A., & Leung, P. (2011). Assessment for learning in language classrooms. In R. Berry and B. Adamson (Eds.), *Assessment reform in education, education in the Asia-Pacific region: Issues, concerns and prospects 14* (pp. 135-154). Netherlands: Springer. doi: 10.1007/978-94-007-0729-0_10
- [12] Cohen, J. W. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- [13] Council of Europe (CoE). (2001). Common European framework of reference for languages: Learning, teaching, assessment. Cambridge, England: Cambridge University Press.
- [14] Darling-Hammond, L. (2010). Performance counts: Assessment systems that support high-quality learning. Council of Chief State School Officers (CCSSO). Retrieved on 12 September 2013 from http://www.ccsso.org/Documents/2010/Performance_Counts_Assessment_Systems_2010.pdf
- [15] Davison, C. (2007). Views from the chalkface: School-based assessment in Hong Kong. *Language Assessment Quarterly*, 4(1), 37-68. doi: 10.1080/15434300701348359.
- [16] Davison, C., & Hamp-Lyons, L. (2010) The Hong Kong Certificate of Education: School-based assessment reform in Hong Kong English language education. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 248-264). New York, NY: Routledge.
- [17] Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393-415.
- [18] Earl, L., & Katz, S. (2006). Rethinking classroom assessment with purpose in mind. Winnipeg, Manitoba: Western Northern Canadian Protocol.
- [19] East, M., & Scott, A. (2011). Assessing the foreign language proficiency of high school students in New Zealand: From the traditional to the innovative. *Language Assessment Quarterly*, 8(2), 179-189. doi: 10.1080/15434303.2010.538779.
- [20] Field, A. (2009). Discovering statistics for SPSS (3rd ed.). Los Angeles, CA: SAGE Publications.
- [21] Harlen, W., & Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. *Language Testing*, 21(3), 390-408.
- [22] Hoy, A. W. (2000). Changes in teacher efficacy during the early years of teaching. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. Session 43:22, Qualitative and Quantitative Approaches to Examining Efficacy in Teaching and Learning.
- [23] Kayi-Aydar, H. (2013). Scaffolding language learning in an academic ESL classroom. *ELT Journal*, 67(3), 324-335.
- [24] Kennedy, K. J., Chan, K. S. J., Yu, W. M., & Fok, P. K. (2006). Assessment for productive learning: Forms of assessment and their potential for enhancing learning. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore. Retrieved on 12 September 2013 from http://www.ied.edu.hk/fpece_project/QEF/Download%20area/1_Assess%20for%20productive%20learning%20%28Kerry%20paper%29.pdf.
- [25] Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *The Canadian Modern Language Review*, 64(1), 199-213.
- [26] Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50. doi: <http://dx.doi.org/10.1016/j.jslw.2012.11.003>.
- [27] Mackey, A., & Gass, S. M. (2005). Second language research: Methodology and design. Mahwah, NJ: Lawrence Erlbaum.
- [28] Malveaux, J. (2010). Employee benefits: Examining employee benefits in the midst of a recovering economy. The Society for Human Resource Management (SHRM), SHRM Survey Findings. Retrieved on 30 August 2013 from <http://www.shrm.org/research/surveyfindings/articles/documents/10-0280%20employee%20benefits%20survey%20report-fnl.pdf>.
- [29] Manfreda, K. L., Bosnjak, M., Berzelak, J., Hass, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rate. *International Journal of Market Research*, 50(1), 79-104.
- [30] MoNE. (2013). English language curriculum: Grades 2 to 8. Ankara: MoNE.

- [31] OECD. (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris: OECD/CERI. Retrieved on 12 September 2013 from <http://www.oecd.org/edu/ceri/35661078.pdf>.
- [32] Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications.
- [33] Papadaki, V., & Papadaki, E. (2006). Job satisfaction in social services in Crete, Greece: Social workers' views. *European Journal of Social Work*, 9(4), 479-495.
- [34] Pat-El, R. J., Tillema, H., & Segers, M. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology*, 83(1), 98-113. doi: 10.1111/j.2044-8279.2011.02057.x
- [35] Phakiti, A., & Roever, C. (2011). Current issues and trends in language assessment in Australia and New Zealand. *Language Assessment Quarterly*, 8(2), 103-107. doi: 10.1080/15434303.2011.566397
- [36] Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4-11.
- [37] Rea-Dickins, P. (2007). Classroom-based assessment: Possibilities and pitfalls. In J. Cummins and C. Davison (Eds.), *International handbook of English language teaching: Part one* (pp. 505-520). New York, NY: Springer.
- [38] Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Harlow: Pearson.
- [39] Sardareh, S. A., & Saad, M. R. M. (2012). A sociocultural perspective on assessment for learning: The case of a Malaysian primary school ESL context. *Procedia - Social and Behavioral Sciences*, 66, 343-353. doi: <http://dx.doi.org/10.1016/j.sbspro.2012.11.277>.
- [40] Shih, T., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods*, 20(3), 249-271.
- [41] Stanford, J. (2008). Employee benefits. *The International Public Sector Accounting Standards Board (IPSASB). IPSAS 25*, 1-88. New York. Retrieved on 6 October 2013 from: <http://www.ifac.org/public-sector/projects/employee-benefits>.
- [42] Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328. Retrieved on 12 October 2013 from <http://www.artfulassessment.org/wp-content/uploads/2009/02/fromformativeassessment.pdf>.
- [43] Stiggins, R. (2007). Five assessment myths and their consequences. *Education Week*, 27(8), 28-29.
- [44] Stiggins, R. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*. Portland, OR: ETS Assessment Training Institute.
- [45] Wu, Y. (2013, July). Bridging assessment for learning to self-regulation in Chinese tertiary EFL writing classrooms. Paper presented at the 35th Language Testing and Research Colloquium, Seoul, Korea. Abstract retrieved on 6 October 2013 from <http://www.ltrc2013.or.kr/download/LTRC2013Program0729.pdf>.
- [46] Zhou, J. (2013, July). Reconciliation between assessment for learning and assessment of learning in Chinese award-winning teachers' EFL classrooms. Paper presented at the 35th Language Testing and Research Colloquium, Seoul, Korea. Abstract retrieved on 6 October 2013 from <http://www.ltrc2013.or.kr/download/LTRC2013Program0729.pdf>.

Hüseyin Öz is assistant professor of English language teaching (ELT) at Hacettepe University in Ankara, Turkey. He holds a BA degree in ELT from Hacettepe University, an MA degree in ELT from Middle East Technical University, and a PhD degree in Applied Linguistics from Hacettepe University. He teaches undergraduate and graduate courses in English as a foreign language (EFL) skills, language teaching methods, language assessment, contrastive grammar, and instructional technologies. His areas of interest include language teacher education, language assessment, language teaching methodology, teacher cognition, metacognition, CALL, and curriculum and instruction.