



| Research Article / Araştırma Makalesi |

The Interplay between Assessment Conceptions and Practices of EFL Instructors: A Mixed-method Study

İngilizce Öğretim Görevlilerinin Değerlendirme Kavrayışları ve Uygulamaları Arasındaki İlişki: Bir Karma Desen Araştırması

İrem Aksu Çekiç¹, Müzeyyen Nazlı Güngör²

Keywords

- 1.Assessment conceptions
- 2.Assessment practices
- 3.English language assessment
- 4.EFL instructors

Anahtar Kelimeler

- 1.Değerlendirme kavrayışları
- 2.Değerlendirme uygulamaları
- 3.İngilizce değerlendirme
- 4.İngilizce öğretim görevlileri

Received/Başvuru Tarihi
24.11.2024

Accepted / Kabul Tarihi
25.07.2025

Abstract

Purpose: Teachers' assessment conceptions and beliefs are influential on the assessment practices teachers apply. In that sense, this study aims to explore the interplay between the assessment conceptions of EFL instructors and their self-reported assessment practices.

Design/Methodology/Approach: Data came from 101 EFL instructors working at preparatory schools in Ankara/Türkiye. Quantitative data were collected using the TCoA-III, developed by Brown (2006), while qualitative data were gathered through three semi-structured focus group interview sessions. Descriptive statistics were used to analyse the quantitative data. The qualitative part was analysed through coding by placing the codes under the four pre-determined themes of Brown, which are improvement, school accountability, student accountability, and irrelevance by using MAXQDA. Besides, a self-reflection journal was kept by the researchers to reflect on their opinions and decisions during the data collection and analysis stages.

Findings: The findings of the study pointed that improvement and irrelevance conceptions were the most agreed ones and assessment practices of the EFL instructors in Türkiye vary from Afl tools to AoL tools. The study highlighted some significant implications for assessment component at pre-service curricula, assessment needs of language teachers, and teacher education assessment policy.

Highlights: Instructors are in favour of using a variety of assessment tools. Formative assessment increases student motivation and achievement, which might be the reasons for their favouring Afl tools. Assessment component fails to bridge the theory and practice in pre-service teacher education. This could be solved by making PSTs gain more real-life experiences.

Öz

Çalışmanın amacı: Öğretmenlerin değerlendirme kavrayışları ve inançları, uyguladıkları değerlendirme yöntemleri üzerinde etkili olmaktadır. Bu bağlamda, bu çalışma, İngilizceyi yabancı dil olarak öğreten öğretmenlerin değerlendirme kavrayışları ile kendi bildirdikleri değerlendirme uygulamaları arasındaki etkileşimi incelemeyi amaçlamaktadır.

Materyal ve Yöntem: Veriler, Ankara/Türkiye'de hazırlık okullarında görev yapan 101 İngilizceyi yabancı dil olarak öğreten eğitmenlerden elde edilmiştir. Nicel veriler, Brown (2006) tarafından geliştirilen TCoA-III aracılığıyla toplanırken, nitel veriler üç yarı yapılandırılmış odak grup görüşmesi oturumu yoluyla elde edilmiştir. Çalışmanın nicel kısmını analiz etmek için betimsel istatistikler kullanılmıştır. Nitel kısım ise, MAXQDA yazılımı kullanılarak kodların Brown'un önceden belirlenmiş dört teması (gelişim, okul hesap verebilirliği, öğrenci hesap verebilirliği ve alakasızlık) altında sınıflandırılması yoluyla içerik analiziyle incelenmiştir. Ayrıca, araştırmacılar, veri toplama ve analiz aşamalarında kendi görüş ve kararlarını yansıtmak amacıyla bir öz değerlendirme günlüğü tutmuştur.

Bulgular: Çalışmanın bulguları, gelişim ve alakasızlık kavrayışlarının en çok kabul gören kavrayışlar olduğunu ve Türkiye'deki İngilizceyi yabancı dil olarak öğreten eğitmenlerin değerlendirme uygulamalarının Öğrenme için Değerlendirme (Afl) araçlarından Öğrenmenin Değerlendirilmesi (AoL) araçlarına kadar çeşitlilik gösterdiğini ortaya koymuştur. Çalışma, hizmet öncesi müfredatlarda değerlendirme bileşeni, dil öğretmenlerinin değerlendirme ihtiyaçları ve öğretmen eğitimi değerlendirme politikası konularında önemli çıkarımları vurgulamıştır.

Önemli Vurgular: Eğitmenler, çeşitli değerlendirme araçları kullanmaktan yana bir tutum sergilemektedir. Biçimlendirici değerlendirmenin öğrenci motivasyonunu ve başarısını artırdığı, bu durumun eğitmenlerin Öğrenme için Değerlendirme (Afl) araçlarını tercih etmesinin nedenlerinden biri olabileceği düşünülmektedir. Ancak hizmet öncesi öğretmen eğitimi programlarındaki değerlendirme bileşeni, teori ile uygulama arasındaki bağı kurmada yetersiz kalmaktadır. Bu sorun, öğretmen adaylarının daha fazla gerçek hayat deneyimi kazanmalarının sağlanmasıyla çözülebilir.

¹İzmir University of Economics, School of Foreign Languages, İzmir, Türkiye; <https://orcid.org/0009-0000-2678-8922>

²Corresponding author, Gazi University, Gazi Faculty of Education, Foreign Languages Education Department, ELT Programme, Ankara, Türkiye; <https://orcid.org/0009-0000-2678-8922>

INTRODUCTION

Recent studies on teacher education are concerned with how teachers' histories, beliefs, and conceptions are formed and established, and how knowledge about teaching, learning, and assessment affect their decision-making and classroom practices (Lutovac & Flores, 2022). Conceptions of assessment refer to "one's beliefs, meanings, and understandings of assessment" (Fletcher et al., 2012, p. 120) in this study. Besides, assessment practices are wide and categorised differently in the assessment literature. For instance, in addition to the categorisations, such as formative, summative assessment and traditional, alternative assessment, the terms Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL) were coined as alternatives for summative and formative assessment (Schellekens et al., 2021). These categorisations demonstrate that teachers may favour using different types of assessment practices based on their conceptions and understanding of assessment and that their assessment practices are reflections of their assessment conceptions (Harris & Brown, 2009).

The assessment practices of teachers are the results of the conceptions they own (Harris & Brown, 2009) and their conceptions are formed during their education in teacher education programmes (Kyttälä et al., 2022). This is why it is of paramount importance to understand whether English Language Teaching programmes (ELTEP) provide pre-service teachers (PSTs) with sufficient education concerning assessment and the way these courses shape their assessment conceptions, which might pave the way to reconsider the nature and content of these courses. Finally, it is also crucial to reach an in-depth comprehension of teachers' opinions on assessment practices, which may result in making them more effective institutionally. Besides, in-service teacher education can be prepared considering their assessment conceptions and competencies in that regard.

Considering the importance of assessment as one of the pillars of teaching and learning and teachers' active role in the process, the study aims to explore the interplay between the assessment conceptions of Turkish EFL instructors who work at preparatory schools at tertiary level and teacher assessment practices through the four-facet model developed by Brown (2002, 2004, 2008), and classroom practices to understand the relation between each of these dynamics. Accordingly, the research questions are as follows:

1. What are the assessment conceptions of Turkish EFL instructors in terms of four components defined by Brown (2002, 2004, 2008)?
2. What are the assessment practices that Turkish EFL instructors apply in their institutions?
3. What is the relationship between Turkish EFL instructors' assessment conceptions and the assessment practices they apply?

Theoretical framework

Assessment Conceptions of EFL Teachers

Assessment conceptions are about the beliefs and comprehension teachers have concerning assessment (Lutovac & Flores, 2022). Studies on assessment conceptions revealed many features of assessment conceptions. First, teachers may hold more than one conception, which may even conflict with each other. For instance, Azis (2015) indicates that the conflicting conceptions of teachers are due to their willingness to use process-oriented methods and the requirements of the national exam policy. Second, teachers' experiences as professionals, as well as their own experiences as students may affect their conceptions (Kyttälä et al., 2022). Finally, the influences of culture and the education system on teachers' assessment conceptions are also argued. For instance, Xu and He (2019), whose study led the participants to reflect on their assessment practices, indicated that school assessment culture and already-existing rooted assessment routines, such as being an exam-dominated country, may set limitations on their conception change.

Brown (2002, 2004, 2008) has conducted many studies on assessment conceptions of teachers by proposing a four-facet model and scales to understand assessment conceptions of teachers. Specifically, the Teachers' Conceptions of Assessment Abridged Scale (TCoA-IIIa) (Brown, 2006) was adopted in this study since it was developed later than the full form of the scale and can reach similar information. The four subscales/factors defined by Brown (2002, 2004, 2008) are improvement, school accountability, student accountability, and irrelevance.

Firstly, assessment is seen as positive and constructive in improvement conception since it indicates that assessment is about facilitating student learning with the necessary changes in teaching and providing feedback. Therefore, assessment needs to provide the best information about students' knowledge and be valid and reliable (Brown, 2008; Monteiro et al., 2021). To this end, a variety of assessment tools are used by the teachers having this conception.

Secondly, assessment is seen as a way to demonstrate the success of the teacher, the school, and the country in school accountability conception. Hence, the focus is on the results of assessment (Brown, 2008). Historically, school accountability was valid since students' success was the determinant of the budget provided to schools (William, 2010). Another rationale for school accountability conception is to make positive changes in the teaching quality by ensuring quality institutions and making changes in the teachers' opinions, if necessary.

Thirdly, student accountability conception is about holding students responsible for their learning by checking their performances using a criterion, grading and placing them into different levels/classes, and comparing students' performance with other students (Brown, 2004). Holding students accountable for their learning and performance has some rationales, such as

improving student motivation and penalising the irresponsible ones, as indicated by Kahn (2000). Besides, the popularity of student-centred methods, such as peer assessment and self-assessment also paves the way for student accountability conception (Brown, 2008).

The final conception, irrelevance, points out that teachers' judgements and observations of students are sufficient without formal assessment methods. There might be many reasons for holding this conception. For instance, some teachers may believe that assessment results are ignored, or they do not reflect students' real performance (Brown, 2004). Moreover, the fact that assessments lead to be focused on exams rather than the learning process is another concern, which Berry (2008) calls the negative washback effect. Some of the other reasons are related to perceiving assessment as equivalent to testing, leading to negative emotions (Brown, 2008).

In-service EFL teachers' assessment practices

Teachers' skills and knowledge are highly underlined in the assessment literature as they are responsible for many things from grading exam papers to making decisions on actions that will lead to student success. Therefore, investigating teachers' conceptions and how they put these conceptions into practice is crucial. Supporting this, Popham (2011) points out the positive correlation between teachers' attributing significance to assessment and application of a variety of meaningful classroom practices. Furthermore, the quality of education provided is linked to the quality of teachers (Lutovac & Flores, 2022), suggesting the necessity to consider their educational background since their assessment conceptions and competencies have begun to develop in their pre-service years (Kyttälä et al., 2022). In the Turkish context, English teachers study for four years in ELTEP following a nationwide curriculum, which includes only one course about testing and evaluation throughout the eight-term in a four-year period, which is not sufficient (Güngör & Güngör, 2024). In this vein, teachers' feeling incompetent in assessment has been indicated in some studies, which argue that this lack of competency might be about the content, nature, or delivery of the courses they receive as PSTs. This affects their in-service assessment practices negatively. Moreover, they face some other problems when they start their profession, such as syllabus and time constraints, which limit teachers in putting their assessment literacy into practice (Ogan-Bekiroglu & Suzuk, 2014). In language classes, assessment can occur in distinct forms. Firstly, one of the most popular categorisations is summative and formative assessments. While summative assessment includes scoring and can be seen in the form of final examinations and university entrance examinations, formative assessment is concerned with the learners' progress and efforts to improve the learning process through self-assessment, peer assessment, and portfolios (Schellekens et al., 2021). Teachers prefer both of these assessment methods (Gebiril, 2016), while Kyttälä et al. (2022) claim that formative assessment is gaining more importance.

Secondly, traditional and alternative assessments are investigated in the literature. Traditional assessment methods include true-false and multiple-choice exercises, which are far from being authentic (Nasab, 2015). On the other hand, alternative assessment methods are focused on the higher-level thinking skills of students, together with their problem-solving skills by including real-life contexts, as Norris et al. (1998) address. Therefore, portfolios and projects can be considered as examples of this type.

Lastly, the terms AaL, AfL, and AoL were introduced relatively later than the previous categorisations. AfL is described as a part of a routine of students and teachers for the enhancement of learning through dialogues, observations, and demonstrations (Schellekens et al., 2021). Although AaL is sometimes interpreted as a sub-branch of AfL, Lam (2015) highlights the significant role of the learners and their independence in the assessment process in AaL. The last component of this categorisation is AoL which can be considered in a similar framework as summative assessment since the aim of that type is to provide certifications and reach a conclusion about students, which may be seen in the form of exams (Lee, 2011).

Although a bulk of research on beliefs and conceptions of assessment has been carried out in pre-service and in-service teacher education (DeLuca, et al., 2019), studies on understanding teachers' conceptions of assessment and exploring the relation between these conceptions and their assessment practices are scarce. Teachers may begin their pre-service education with varying levels of assessment knowledge and experience; however, their assessment conceptions and practices may change in their in-service years due to personal, professional, and external political contextual factors (Mockler, 2011). In other words, in-service teachers' conceptions may reflect cultural and societal practices and may be context dependent. Therefore, this study aims to understand EFL instructors' assessment conceptions and practices and explore the interplay between their assessment conceptions and assessment-related practices in the Turkish tertiary education context.

METHOD

Research Context

This study was carried out in preparatory schools of six universities in Ankara/Türkiye. A one-year preparatory school programme is offered to students in higher education to improve their skills in English upon the university entrance and proficiency exams. This practice has been employed in Türkiye since 1958, even though there have been different regulations since then (British Council & TEPAV, 2015). Accordingly, students have to take the proficiency exam at the beginning of the school year, which gives an idea about the English level of the students and decides if a student will be exempted from the preparatory school. At the end of the preparatory school year, students are expected to take a proficiency exam again and they have to study in the

preparatory school one more semester or one more year when they fail. They are expected to pass the proficiency exam by the end of two years. EFL instructors are employed at these preparatory schools to teach English courses after proving their proficiency in the English language (Council of Higher Education, 2016).

Participants

The participants of the quantitative part consisted of one hundred one EFL instructors working at six different preparatory schools in Ankara/Türkiye. They had different educational backgrounds and experience years in teaching English. Table 1 presents the demographic information of the participants.

Table 1. Frequency Distributions of the Participants

How long have you been teaching?	Frequency	Per cent
Less than five years	21	20.8
Between 5-10 years	15	14.9
More than ten years	65	64.4

Which department did you graduate from?	Frequency	Per cent
English Language Teaching	68	67.3
English Language & Literature	23	22.8
Other	10	9.9
Total	101	100.0

Following the quantitative part, nine participants voluntarily agreed to take part in the focus group interview sessions (Table 2).

Table 2. Participants of the Focus Group Interview Part

Participants (Pseudonyms)	Session Number	Years of Experience	Undergraduate Department	Instructor at
Elif	Session 1 (I1)	More than ten years	English Language Teaching	University A
Esra	Session 1 (I1)	Between 5-10 years	English Language & Literature	University A
Leyla	Session 1 (I1)	Less than five years	English Language Teaching	University B
Ekin	Session 2 (I2)	Between 5-10 years	Translation & Interpretation	University A
Melda	Session 2 (I2)	More than ten years	English Language Teaching	University C
Ebru	Session 2 (I2)	More than ten years	English Language Teaching	University D
Hatice	Session 3 (I3)	Less than five years	English Language Teaching	University E
Buket	Session 3 (I3)	More than ten years	English Language Teaching	University C
Songül	Session 3 (I3)	Less than five years	English Language & Literature	University F

Data Collection and Analysis

This study followed a mixed-method research design. The quantitative data came from the 27-item Teachers' Conceptions of Assessment Abridged Scale (TCoA-IIIA) developed by Brown (2006). It was administered to one hundred and one EFL instructors working at six different preparatory schools in Ankara/Türkiye, who were chosen through convenience sampling. The qualitative data was gathered through a focus group interview. It was conducted after the pilot study to gather a more comprehensive understanding of the assessment conceptions of the participants and their opinions on assessment practices. Because the pilot study lasted approximately one hour with two participants, the researchers decided to conduct focus group interviews in three different sessions with three participants in each session. This would lead to gathering of more comprehensive data and lessen the time constraints that would be faced due to including nine participants in one session. The duration of the interview sessions was 72, 59, and 65 minutes, respectively. Finally, a self-reflection journal was kept during the data collection and analysis stages to improve the coding process, ensure triangulation, and present the opinions of the researchers to the readers in an explicit way. One entry from the researcher's self-reflection journal, which was written after the pilot study, is presented below.

"I am happy that my friends could express their opinions without hesitation; however, I have other concerns now. This interview took almost one hour with two participants, so how am I going to include 7-9 participants in the same interview, and

how long will it take? Including a small number of participants may make the discussion more effective [...]” (Aksu Çekiç, 2024, p. 39).

As for the data analysis methods, quantitative data analysis was completed through descriptive statistics. Qualitative data, on the other hand, was analysed through coding (Creswell, 2007). First, the data was transcribed verbatim. Second, some repetitive statements of the participants were coded through MAXQDA, and these codes were placed and explained under the themes of the four-facet model of Brown (2002, 2004, 2008), whose scale was already used in the first step of the study.

Validity and reliability

As the quantitative part of the study was collected via TCoA-IIIA, validity and reliability analysis of the scale was run first. Confirmatory Factor Analysis (CFA) was used to check the validity of the scale. The results revealed that the model was structurally valid, and our data was suitable for the model and the four subscales (fit indices: $\chi^2/df = 2.30 < 5$). However, it should be noted that RMSEA, SRMR, and CFI values were 0.114, 0.136, and 0.680 in the respective order, which was due to the limited sample size and this limitation was noted.

Reliability analysis of the scale was carried out with Cronbach’s Alpha to understand the internal consistency of the scale. The results demonstrated that the overall reliability of the TCoA-IIIA was 0.863, indicating a good internal consistency. The reliability coefficient of the four subscales was also satisfactory with the following scores: 0.790 for school accountability, 0.603 for student accountability, 0.900 for improvement, and 0.749 for irrelevance.

Credibility and trustworthiness

To achieve credibility and trustworthiness, peer review, triangulation, prolonged engagement, and member checking (Creswell, 2007) were ensured in the data collection and analysis phases. First, one of the researchers generated the codes during the data analysis stage. Subsequently, these codes were reviewed by the other researcher to ensure peer review. Second, triangulation was achieved by integrating different data collection tools to have a deeper understanding of the topic. In other words, the researchers integrated a scale, a focus group interview, along with their self-reflection journal. Moreover, prolonged engagement was valid since the researchers had similar backgrounds as the participants and one of them was a member of the investigated community. Lastly, after the coding process ended, one of the focus group participants were asked to check the codes to eliminate any confusion or misunderstandings and ensure credibility and trustworthiness.

FINDINGS AND DISCUSSION

The assessment conceptions of EFL instructors in terms of Brown’s four components

The first research question aimed to understand the assessment conceptions of EFL instructors in terms of Brown’s four components. Descriptive statistics (Table 3) show that the highest mean score belongs to improvement conception ($M=50.73$) and that it is followed by irrelevance conception ($M=30.13$). While student accountability conception has the third highest mean score and agreement level among the participants ($M=12.64$), the mean score of school accountability score is slightly lower with the lowest agreement level ($M=11.55$).

Table 3. Descriptive Statistics Results

Subscales	Minimum	Maximum	Mean	Std. Deviation
Student Acc. Score	6	18	12.64	2.496
School Acc. Score	4	18	11.55	2.837
Irrelevance Score	14	43	30.13	6.876
Improvement Score	22	70	50.73	9.106

Even though descriptive statistics gave a summary of the data gathered via TCoA-IIIA, it was necessary to delve into the opinions of participants on assessment as well as assessment practices. Hence, interview data were analysed and interpreted according to the four factors defined by Brown (2002, 2004, 2008) as predetermined themes (Table 4).

Table 4. Pre-Determined Themes and Codes

Improvement	Irrelevance	Student Accountability	School Accountability
Essentiality	Exams as Inaccurate Indicators of Student Performance	Including Peer Assessment and Self-Assessment	No Direct Link Between Institutional/Teacher Success and Student Success
Assistance for Teachers	National Contextual Factors	Guide for Students	Solutions to Diminish the Feeling of Incompetence
Guidance	Teacher Competencies	Proficiency Exam as a High-Stakes Exam	Washback Effect
Meticulous Design	Ineffectiveness of Pre-Service Assessment Courses	High-Stakes Assessment and Student Emotions	
	Assessment Courses in Shaping Pre-service Teachers’ Mindset		
	Standardisation Problems		

The mean scores of the four subscales indicate that teachers mainly hold improvement conception ($M=50.73$), which is followed by irrelevance conception (30.13). While school accountability conception (11.55) is the least agreed upon by the participants, the difference between student accountability ($M=12.64$) and school accountability is not significant. It was seen that focus group interview sessions supported the findings of the descriptive statistics results. The differences between assessment conceptions depend on many different personal, professional and political factors. Firstly, similar to many other assessment conception studies (Ma & Bui, 2021), the participants in this study mainly have the improvement conception, which indicates that assessment is perceived as a positive factor in improving teaching and learning by reflecting on the existing teaching practices with the help of assessment results. Irrelevance is one of the conceptions with the highest mean score in this study, unlike many other studies (Brown, 2002). These results as well as participants' comments indicated that although teachers believe the significance of assessment based on their personal experiences, some professional or political factors, such as the education system in Türkiye, lead them to have question marks about the quality of assessment. Student accountability conception has the third place in this study, which may relate to the political factors and the fact that students take the English proficiency exam at the end of the year. The final conception in this study is school accountability, following student accountability conception. It demonstrates that even though teachers are aware of the importance of assessment and try to apply it effectively in their classes, their main motivation is not demonstrating their own success or their students' success publicly. These conceptions were explained and exemplified below in order of significance of the findings demonstrating how personal, political, and professional factors affect instructors' opinions and conceptions. Kytälä et al. (2022) describe personal context as teachers' experiences regarding assessment outside of teacher education, while professional factors are described as their experiences throughout their teacher education programme as well as their experiences as teachers. Finally, political factors include the political and societal decisions affecting pre-service teachers' conceptions based on what they are expected to do.

Improvement

Descriptive statistics results revealed that the most agreed assessment conception was improvement. This shows that EFL instructors working at the preparatory schools in the Turkish context are quite aware that assessment holds a significant place in teaching and learning, which might be based on personal or professional factors. In other words, professional factors, such as the theoretical part of their undergraduate courses might have convinced them to think that assessment leads to positive results, although they criticised the practical part of those courses. Supporting this, the researchers identified four beliefs, as revealed by codes within the scope of improvement conception. First of all, the participants believe that assessment is an important component in the language learning process. They also think assessment results help them make decisions about their students and students' needs. Some examples of these beliefs are as below.

For me, assessment is useful for getting feedback from students, especially because you know what is conveyed and what is not conveyed to the students after you assess something. [...] (Essentiality, Ebru-12)

[...] Assessment not only assesses learners' mechanic knowledge of English but also it should assess pedagogically; maybe learners' needs and how they react to learning a new language in a classroom atmosphere because there are many individual differences when we observe a classroom... So, it not only assesses learners' exam results but also assesses their psychological needs. (Assistance for Teachers, Hatice-13)

Moreover, participants provided explanations and examples of how they made use of assessment results, indicating that assessment results are a guide for them. Also, they addressed the issue of preparing assessments carefully to attain meaningful results from the assessment, which was indicated by Brown (2008) while explaining improvement conception.

I have an example of that. Once, most of the students in my class were not able to get the difference between "mustn't" and "don't have to." Then, I realised that I must have done something wrong while I was teaching that part, and I went over that one more time for my students. The exam results helped me to understand this. So, of course, we use [assessment] to reflect on our teaching and students' learning. (Guidance, Buket-13)

Improvement conceptions holding the first place is parallel to the findings of other studies (Pastore & Mincu, 2024) indicating that teachers who work in different contexts, such as the Chinese, Italian, and Turkish contexts relate assessment with improvement regardless of the level that they teach. In other words, the findings of the present study support the other assessment conception studies in which teachers working at primary, secondary and high schools are investigated. This means that there is consistency in assessment culture across different contexts and levels.

Irrelevance

Irrelevance conception had the second highest mean score following improvement conception. It should be underlined that teachers can hold different or conflicting conceptions at the same time (Lutovac & Flores, 2022), which may explain improvement and irrelevance conceptions' having the highest mean scores in this study.

Focus group interview sessions paved the way for us to understand the underlying reasons. First, many participants stated their concerns about the quality of exams. They believed exams were irrelevant and not true indicators of students' performance due to different problems. First, there are problems concerning the preparation stage of the exams. Besides, implementing decontextualised exams, and students' being anxious as a result of taking exams frequently are some of the concerns mentioned by the participants. Additionally, the participants stated that traditional methods of assessment are frequently used in the Turkish education system. They illustrated this issue by providing examples from their personal experiences as students, as well as their

current observations. Buket exemplified her experiences in the third interview, which demonstrated that just having multiple choice exams did not improve her English skills as a high school student.

When I first entered my university, I had 99 correct answers out of 100, and I was not able to speak English. My instructor asked me to introduce myself, and I said I could not speak English, and I thought [it was] okay. As a picture, I was a perfect student, but when I wanted to speak English, I could not. [...] (Exams as Inaccurate Indicators of Student Performance, Buket, I3)

In our Turkish context, assessment is generally seen as something done by the professor or by the teacher, but we do have peer assessment and self-assessment, right? So, we could look at how assessment impacts their learning, maybe from this perspective as well. Generally, [...] if something is not in the test, students generally do not study and pay attention to it. [...] (National Contextual Factors, Melda, I2)

This conflict between the overreliance on the traditional exercises in different countries' education systems and teachers' opposite perspectives is also underlined by Azis (2015). He claims that there is a clash between teachers' willingness to apply process-oriented methods and national assessment policies. Similarly, Tsagari and Vogt (2017) state that the policies of a nation have an impact on teachers' assessment practices. Traces of the Turkish education system could be detected from the participants' words, which exemplifies how political factors might shape or conflict with the assessment conceptions of teachers. For instance, Melda (I2) states that even though she prefers self-assessment or peer assessment, her institution has different policies, which she perceives as the reflection of the assessment perception in Türkiye. Moreover, the criticisms of Buket can be seen below.

[...] I think we lack assessing interaction in most preparatory schools, not only in our school, because we think this is very difficult. [...] So, an addition of paired assessment, group assessment, and assessing the interaction is something that needs to be added in most of the preparatory schools in Türkiye, I think. (National Contextual Factors, Buket-I3)

Moreover, professional factors, such as insufficient pre-service education of teachers, may lead them to think that assessment is irrelevant (Vardar, 2010). In this vein, it is worth noting that although the participants of this study did not consider assessment irrelevant, they had question marks about their assessment competencies. While some of the instructors mentioned that they had to improve themselves concerning in-class assessment, some, on the other hand, focused on the exam preparation part as illustrated below.

[...] I am not good at writing distractors. This is a confession; this is sad, but it is what it is. I am not good at writing distractors, especially in reading. (Teacher Competencies, Ekin-I2)

I feel that I need to learn effective ways of assessing the language skills of our students. (Teacher Competencies, Esra-I1)

Participants' reflections on their competencies concerning assessment addressed the need for in-service training or continuous development programmes in the Turkish context supporting the statements of Larenas and Brunfaut (2023), who highlight the lack of confidence among language teachers concerning assessment. In this regard, the participants criticised the one-term assessment courses in pre-service education since it was not enough to prepare them for real-life contexts, and the content of this assessment-related course/s was not effective in providing them with sufficient training on assessment. Besides, they believed that the course/s could not depict a true picture of what assessment was while shaping their mindset concerning the topic. Similarly, Mertler argues (2003) that teachers' practical experiences are more helpful for them compared to their pre-service education on assessment. The reflections of the researcher written in her self-reflection journal and the participants' words are provided below.

[...] They definitely needed more knowledge when it came to assessment. However, what they really need is practical information about real life, not theories or names, such as formative or summative assessment etc.... What I think is arranging some seminars in relation to real practices, as well as cooperation between teachers would be better [...] (Aksu Çekiç, 2024, pp. 64-65).

[...] Well, I always criticised my university's approach. I learned a lot of things, wrote a lot of research papers, read a lot of articles, etc. [I remember] all the terminology and methodology about teaching, but it is not about real life. The moment I graduated and started working, real life was not like it was given in the books. I always criticise this aspect of my university. But I remember what validity and reliability are about assessment or assessment types: formal assessment, summative assessment... I can list all of them, but in practice, I mean, how to assess during class while a student is speaking, how to assess if he/she is doing well, what aspects should I note down or should I ignore? Nobody told me anything about those. (Ineffectiveness of Pre-Service Assessment Courses, Elif-I1)

Finally, Brown (2008) states the fact that the subjective opinions of teachers are involved while assessing students, which may distort the meaning of the assessment. Concerns about standardisation especially in writing and speaking exams were underlined by the participants of this study in different interview sessions, which is a part of professional and personal factors because their observations as well as their experiences lead teachers to have this conclusion. The participants think that ensuring standardisation while assessing students' language skills, such as writing and speaking skills, is a problem for them. This might lead to these results being evaluated unfairly, paving the way for the irrelevance conception.

[...] The written and the spoken assessments are more likely to be instructors' way of assessing. At that point, I guess there must be a standard for teachers' perspective because I might think of giving a score of 15 for a writing paper, but the same paper can be assessed as 25 [by another instructor.] That is a huge gap between the two instructors, so at that point, there must be a standard, and I do not believe that there is a perfect standardisation for the assessment of writing or spoken exams. [...] (Standardisation Problems, Hatice-I3)

In short, the findings of the study demonstrated participants' concerns about assessment, similar to the ones listed by Brown (2008) while explaining irrelevance conception. These concerns involve exams' not being true indicators of student' performance, their own competencies as teachers, standardisation problems in grading, and the traditional point of view in

assessment, which indicates the adoption of standardised exams. These concerns should be investigated by the institutions and policymakers so that they can be tackled cautiously.

Student accountability

Brown (2008) mentioned that including student-centred methods, such as peer assessment and self-assessment tools for grading purposes is concerned with making students accountable for their learning process. Similarly, the participants mentioned that they utilised peer assessment and self-assessment tools in their assessment practices. Moreover, they believed that students should be reflective and use their assessment results to draw a map for their learning and take the required steps in the learning process as active agents through different practices. Unlike the participants of Tsagari and Vogt's (2017) study, who did not prefer alternative assessment tools, the participants in our study indicated that they preferred AfL practices.

Learners also need to see what they have done and what they could not do on a task, for example, at the end of this assessment. It does not have to be graded or something but in any kind of assessment. That is like a sign for them or a guide for them to see which parts they have learned better or which parts they should improve. [...] They can reflect on their learning if they are clever enough, well, if they are good reflectors, let us say. (Guide for Students, Elif-I1)

Furthermore, students are held accountable for their learning through their performances, and their success is decided based on a criterion or based on other students' performances, leading them to be placed in different classes/schools accordingly, as Brown (2008) explained. This is similar to proficiency exam administration in preparatory schools. Proficiency exams in preparatory schools are the determinants of the students' language levels and they show if students will skip the preparatory school to start their departmental courses. Thus, it is an influential political factor in their conceptions since instructors have to prepare their students for this exam. Some participants mentioned how they used proficiency exam results to put students into different classes. However, they also repeatedly criticised how high-stakes exams, including language proficiency exams, negatively affect students' psychology. Similarly, Harlen and Crick (2003) point out that teacher assessment is less harmful to students' self-efficacy and more motivating for students than high-stakes assessments.

In my institution, we have a slightly different approach. If students cannot reach the goals of a level, we generally use those assessment results and change their level. We kind of demote them to a lower level, and we also use students' successful results and give them another chance to move up one more level [...], so we use assessment results like this in addition to what has been said. (Proficiency Exam as a High-Stakes Exam, Melda-I2)

[...] I think we should make [assessment] a bit more individualistic and softer than what we are doing right now because it makes me sad to see students getting stressed over some useless preposition items in a long phrase that they have to memorise [...] (High-Stakes Assessment and Student Emotions, Songül-I3)

In conclusion, the participants are in favour of holding students accountable through peer assessment and self-assessment methods, and individual guidance. Nevertheless, political factors, such as high-stakes exam context and the observation that students are excessively held accountable for their learning, and they feel under pressure may contribute to teachers' adopting irrelevance conception.

School accountability

School accountability conception suggests that assessment results are a way for schools and teachers to show that they have been successful (Brown, 2008). Therefore, teachers with this conception may point to a direct link between students' success and institutional success or teachers' success. Interview data reveal that participants cannot draw a direct link between student success and teacher/institutional success. However, they hold themselves responsible for student learning and are willing to improve their assessment competencies. Therefore, their opinions and suggestions to overcome the feeling of incompetency were included within the scope of the school accountability conception. Communication and cooperation between teachers were emphasised as significant points in that regard, which is also mentioned by Yan and Fan (2021). The participants in their study also highlighted communication between stakeholders as a neglected point in language testing, which is said to be helpful for them to improve themselves.

I think you can roughly call a teacher successful if a high percentage of students get above a certain grade, but not an institution because institutional practices change quite a lot. [Also,] "successful" in one institution's terms and one institution's assessment criteria might be very unsuccessful in another, but you can say that this teacher has done a good job or a bad job. But I think you have to take care of the fact that certain groups sometimes have certain behavioural issues [...], so you can carefully assess the teacher but not the institution, I think. (No Direct Link Between Institutional/Teacher Success and Student Success, Leyla-I1)

Well, collaboration, I think because we are not alone, as you can see [in the interview.] Everybody has the same problem. I do not want to say, "Okay, let us go and read articles." because I do not find it very practical. However, I think the instructors should come together and do research or share experiences in their institutions. Maybe they can read and research together, especially with the testing and assessment team because we are doing some kind of assessment in the class, and that is fine. (Solutions to Diminish the Feeling of Incompetence, Elif I1)

Brown (2008) mentions that school accountability conception and the emphasis on exams and demonstrating the success of schools/institutions may lead teachers to turn their focus to exam scores and they may begin to "teach to the test", which is coded as the washback effect by the researchers. The participants in this study criticised some of their students as they focused more on the topics and skills that they would encounter in their exams, indicating the negative washback effect, as also acknowledged by Yan et al. (2018). They indicated that the negative washback effect or "teaching to the test" is seen in exam-

oriented countries, such as China where the main purpose of teachers is to prepare their students for the high school entrance exam, which is similar to the Turkish context. Thus, it is of no surprise for students to turn their focus to exams and only study for these exams affecting their future lives.

It may have some positive and negative effects. Well, when I say negative effect, for example, students do not have to pass a speaking test to pass the preparatory school in my institution. Therefore, they want to focus on grammar, vocabulary, and reading more than doing some speaking exercises, [but] it is not learning English. I mean, while learning English, they lack the skill of speaking most of the time just because of that. (Washback Effect, Ebru-12)

In conclusion, the learning environment is sometimes negatively affected by the fact that students' knowledge will be assessed through exams, which may lead students and teachers to focus on exam results, rather than focusing on language learning in general, and it is indicated as a concern by the participants. Furthermore, the findings of the study show that there is not a direct parallelism between student success and teachers'/institutions' success, which means that students' success or failure may depend on many different external factors regardless of how hard teachers or institutions try to fully prepare them for their exams.

Overall, the findings revealed that EFL instructors mainly had improvement conception, which is followed by irrelevance conception since instructors had some concerns about both their competencies and the existing assessment practices in their institutions. The focus was less on the school accountability conception in this study.

EFL instructors' assessment practices in their institutions

The second research question aimed to understand the assessment practices EFL instructors applied in their institutions. The findings are based on participants' self-reported assessment practices. In other words, they were asked the assessment practices they use in their institutions and classes during the interview. The findings revealed that preparatory schools of the six investigated universities employ a variety of assessment practices from AfL to AoL tools. It was found that assessment practices in these six institutions are similar to each other. Ekin gives a detailed description of some assessment practices that take place in her institution. After that, she explains how they deal with the assessment of speaking and writing skills. Furthermore, some in-class assessments are said to be based on teachers' own decisions, rather than their institutions' obligatory practices, although the participants feel like they do have only minor individual choices apart from the institutional assessment policies. Similarly, Troudi et al. (2009), argue that teachers are not involved in the decision-making process of assessment tools. In other words, EFL instructors may not play an active role in the decision-making process of assessment tools or preparation stage of them. Still, they might have some individual choices in class. For example, Esra explains how she gives feedback in class.

Of course, portfolios have some writing and speaking tasks. Writing tasks evolve during the term. For example, [...] we ask questions to describe the process, and it is also a process writing [written in] two drafts. [...] We give feedback, and the student writes and gives us his/her second draft. Then, we give feedback again with his/her grade, and they learn. First, they write something little, and then they write a paragraph. Then, they learn how to write an essay, and from that time, they continue writing several types of essays. In the proficiency exam, they are expected to write a well-developed essay. When it comes to these speaking tasks, they are mostly about real-life issues. They mostly include some role-plays that can be with a partner or alone, and they range from two to four minutes. [...] (Ekin, 12)

In the class, I use peer feedback. For example, they write a paragraph, and I want the students to give feedback to their classmates. [...] I do not use it in speaking because they can be really shy and frustrated, so I do not use that. [...] (Esra, 11)

The participants also indicated that they had some suggestions and criticisms for the existing assessment practices in their institutions to enhance the effectiveness of those practices.

[...] So, some of the practices, unfortunately, are determined by the logistics as well. For example, speaking assessment; we want to do more of it, but we have to do it on a slow scale, not very frequently. This makes the teachers more responsible for evaluating students' speaking skills. [...] (Melda, 12)

I do not think anything should be omitted [from the institutional assessment practices] but added. I think we lack assessing interaction in most preparatory schools, not only in our school, because we think this is very difficult. [...] (Buket, 13)

As a result, it was found that a variety of assessment tools are used in preparatory schools in Ankara/Türkiye, which is similar to the findings of Struyven et al. (2005) who indicate that portfolios, peer assessment and self-assessment tools are applied at tertiary level.

The interplay between EFL instructors' assessment conceptions and their assessment practices

The last research question explored the interplay between EFL instructors' assessment conceptions and assessment practices. Firstly, the findings demonstrated that participants mainly have improvement conception, which is supported by the assessment practices they prefer. The participants are in favour of using a variety of assessment tools, just like mentioned by Troudi et al. (2009). They indicate that different assessment tools should be used so that it can be helpful for students to benefit from assessment. In this vein, it is found that participants in this study use quizzes, peer feedback, games, and group discussions as in-class assessment tools, in line with improvement conception. Besides, the participants also stated that they preferred including AfL tools in their lessons by focusing on students' production, performance, and progress rather than their exam results. As DeLuca et al. (2013) underlined, formative assessment increases student motivation and achievement, which might be the reason for their favouring AfL tools.

Secondly, although the participants of this study did not directly state that assessment is irrelevant, they had some concerns and question marks about the existing assessment practices, which might have resulted in 'irrelevance conception' being the second one in this study. Some criticisms of the participants include traditional assessment methods and inauthentic and decontextualised exams. Teachers' criticisms should be considered so that necessary steps can be taken, and the adoption of irrelevance conception can be decreased. Moreover, some institutional problems, such as the lack of number of teachers and school facilities, negatively influence the effectiveness of assessment practices, as mentioned by the participants, along with some other criticisms and suggestions, which led them to have irrelevance conception to some extent. Furthermore, Brown (2002) indicates that teachers may believe that assessment should be based on teachers' observations rather than standardised assessment tools. While the participants in this study, such as Elif (I1) and Ebru (I2), believe that observations of teachers comprise a significant part of assessment, they still believe that a variety of assessment tools should be adapted to improve student learning.

Thirdly, the participants of this study have student accountability conception, which may, in some aspects, go hand in hand with the improvement conception. Participants' having student accountability conception reflected itself in participants' using peer assessment, self-assessment, and student-centred methods, as well as their willingness to provide students feedback by making them active agents of their learning. Yet, the participants were still aware of the pressure that students feel due to being accountable for their performance in high-stakes exams.

The final conception in this study is school accountability. Since the participants do not mainly prefer showing that they have done a successful job as teachers or demonstrate their institutions' success, they also do not favour high-stakes exams. In other words, the participants have school accountability conception only to a limited extent. Similarly, they do not perceive standardised exams as one of the most significant assessment practices. On the contrary, they prefer including a variety of different assessment tools, which is mainly related to improvement conception.

Overall, this study found that there is a positive correlation between assessment conceptions of the instructors and their preferences concerning assessment practices, which is also claimed by Brown (2008). Even though there is a difference between Brown's research context, which investigated primary and secondary school teachers' assessment conceptions, and this study's context, similar results could still be attained.

CONCLUSION AND RECOMMENDATIONS

This study aimed to explore the interplay between their assessment practices and conceptions regarding personal, professional, and political factors. Implications are suggested for teacher education programmes, effective assessment practices, and transparent assessment policies in language teaching and learning. First, as addressed in instructors' interviews, assessment component fails to bridge the theory and practice in pre-service teacher education (Güngör & Güngör, 2024). This could be solved by making PSTs gain more real-life experiences, as underlined by Güngör and Güngör (2024). Moreover, similar to the suggestion in the study of DeLuca et al. (2013), the participants in this study also shed light on some points to be considered when designing courses for PSTs. These include the necessity of focusing on in-class assessments and giving immediate feedback. Moreover, designing quality exams was a point of concern for the participants as they felt a need to improve themselves in that regard, which needs to be included in the pre-service teacher education. As DeLuca et al. (2013) suggest, when PSTs get the opportunity to improve themselves on assessment, they demonstrate positive results, such as higher levels of perceived confidence and competency.

Second, there is a need to have in-service training and continuous professional development programmes by considering teachers' needs so that improvement conception can be fostered, and irrelevance conception can be decreased (Schellekens et al., 2021). The participants of this study also demonstrate their willingness to participate in these programmes to increase their assessment competency level and to increase the effectiveness of the assessment procedures.

Third, in order to make assessment perceived as less irrelevant, the procedures of designing and implementing high-quality assessments should be dealt with cautiously. In this regard, preparing exams carefully and grading speaking and writing exams in a standardised way should be prioritised so that assessment could be valid and meaningful (British Council & TEPAV, 2015). As the participants indicate, testing units of the preparatory schools have a huge responsibility in these regards, which should be supported by the efforts of teachers.

Fourth, the participants of this study underscored that the primary focus should be including more meaningful and less strict and stressful assessment practices. Furthermore, participants' support for AfL tools manifested itself as supporting peer assessment and self-assessment methods. Moreover, portfolio assessment is favoured by the participants. As indicated by Delett et al. (2008) and supported by our participants, portfolio assessment is helpful in increasing student motivation and encouraging reflection.

Finally, this study found that there are some institutional problems in the Turkish higher education context. Solving institutional problems, such as financial problems or problems regarding the number of instructors, will lead to the effective implementation of assessment practices. Therefore, needs analysis studies can be applied to gain better insights and comprehensive understanding of the problems in different institutions so that solutions can be developed accordingly.

Limitations

This study has some limitations. The researchers intended to reach more participants while collecting quantitative data; however, the number of participants was limited to 101 participants from six preparatory schools. Future studies may reach out to more EFL instructors from various institutions, which will help form a broader perspective of the assessment conceptions of EFL instructors in the Turkish context. These conceptions may be compared and contrasted with those of international EFL/native instructors to reach a sound conclusion. Furthermore, the data gathered regarding the assessment practices of the participants is based on their own reports during the focus group interview sessions. Following studies may include on-site classroom observations to eliminate this limitation, get a better understanding of the topic, and enhance the credibility of the study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, author-ship, and/or publication of this article.

Statements of publication ethics

We hereby declare that the study has not unethical issues and that research and publication ethics have been observed carefully.

Author contribution statements

The first and second authors conceived of the presented idea. The first author developed the theory and performed the computations. The second author verified the analytical methods. The second author encouraged the first author to investigate mixed-method research design and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Researchers' contribution rate

The study was conducted and reported with equal collaboration of the researchers.

Ethics Committee Approval Information

This study was conducted for the first author's master's thesis at Gazi University in Ankara/Türkiye from May 2023 to July 2024 with the approval of Ethics Commission at Gazi University under the supervision of the second author.

REFERENCES

- Aksu Çekiş, İ. (2024). *An investigation into the assessment conceptions of in-service English language instructors* (Publication No. 873387) [Master's thesis, Gazi University] <https://tez.yok.gov.tr>
- Azis, A. (2015). Conceptions and practices of assessment: A case of teachers representing improvement conception. *TEFLIN Journal*, 26(2), 129-154.
- Berry, R. (2008). *Assessment for learning*. Hong Kong: Hong Kong University.
- British Council & TEPAV (2015). *The State of English in Higher Education in Turkey*. Ankara. Retrieved from <https://www.britishcouncil.org.tr/>
- Brown, G. T. L. (2002). *Teachers' conceptions of assessment* (Doctoral dissertation). Retrieved from <https://researchspace.auckland.ac.nz/handle/2292/63>
- Brown, G. T. L. (2004) Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99(1), 166-170.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science.
- Council of Higher Education (2016). [Yükseköğretim Kurumlarında Yabancı Dil Öğretimi Ve Yabancı Dille Öğretim Yapılmasında Uyulacak Esaslara İlişkin Yönetmelik]. Official Gazette, Issue: 29662.
- Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches*. The United States of America: SAGE.
- Delett, J. S., Barnhardt, S., & Kevorkian, J. A. (2008). A framework for portfolio assessment in the foreign language classroom. *Foreign Language Annals*, 34(6), 559-568.
- DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator*, 48(2), 128-142.
- DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation*, 61, 159-169.
- Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P., & Rees M. (2012). Faculty and students conceptions of assessment in higher education. *Higher Education*, 64, 119-133.

- Gebril, A. (2016). Language teachers' conceptions of assessment: An Egyptian perspective, *Teacher Development*, 21(1), 81-100.
- Güngör, M. A., & Güngör, M. N. (2024). Relocating assessment in pre-service teacher education: an emerging model from activity theory lens. *European Journal of Teacher Education*, 47(2), 348-368.
- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169-207.
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365-381.
- Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research*, 93(5), 276-286.
- Kyttälä, M., Björn, P. M., Rantamäki, M., Lehesvuori, S., Närhi, V., Aro, M., & Lerkkanen, M. K. (2022). Assessment conceptions of Finnish pre-service teachers. *European Journal of Teacher Education*, 1-19.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.
- Larenas, S. V., & Brunfaut T. (2023). But who trains the language teacher educator who trains the language teacher? An empirical investigation of Chilean EFL teacher educators' language assessment literacy. *Language Testing*, 40(3), 463-492.
- Lee, I. (2011). Bringing innovation to EFL writing through a focus on assessment for learning. *Innovation in Language Learning and Teaching*, 5(1), 19-33.
- Lutovac, S., & Flores, M. A. (2022). Conceptions of assessment in pre-service teachers' narratives of students' failure. *Cambridge Journal of Education*, 52(1), 55-71.
- Ma, M., & Bui, G. (2021). Chinese secondary school teachers' conceptions of L2 assessment: A mixed-methods study. *Studies in Second Language Learning and Teaching*, 11(3), 445- 472.
- Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus. Retrieved from <https://eric.ed.gov/?id=ED482277>
- Mockler, N. (2011). Beyond 'what works': Understanding teacher identity as a practical and political tool. *Teachers and Teaching: Theory and Practice*. Retrieved from <https://www.researchgate.net/>
- Monteiro, V., Mata, L., & Santos, N. N. (2021). Assessment conceptions and practices: Perspectives of primary school teachers and students. *Frontiers in Education*, 6, 1-15.
- Nasab, F. G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, 2(6), 165-178.
- Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments*. University Of Hawaii.
- Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *The Curriculum Journal*, 25(3), 344-371.
- Pastore, S., & Mincu, M. (2024). A feasible balance? The Italian teachers' standpoint on assessment literacy, assessment practice, and teacher professional development. *European Journal of Teacher Education*, 1-20.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46, 265-273.
- Schellekens, L. H., Bok, H. G. J., Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of assessment as learning (AaL), assessment for learning (AfL), and assessment of learning (AoL). *Studies in Educational Evaluation*, 71, 1-15.
- Struyven, K., Dochy F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.
- Troudi, S., Coombe, C., Al-Hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555.
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Vardar, E. (2010). *Sixth, seventh and eighth grade teachers' conception of assessment* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/>
- Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45(2), 107-122.
- Xu, Y., & He, L. (2019). How pre-service teachers' conceptions of assessment change over practicum: Implications for teacher assessment literacy. *Frontiers in Education*, 4, 1-16.
- Yan, X., Zhang, C., & Fan, J. J. (2018). Assessment knowledge is important, but...: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168.
- Yan, X., & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219-246.