Language Tests and Assessment Practices of English Language Teachers through the Lens of EFL Students

Mine Yıldız, Department of English Language Teaching, Atatürk University, Erzurum, Türkiye, mine.yazici@atauni.edu.tr

Recommended Citations:

APA

Yıldız, M. (2024). Language Tests and Assessment Practices of English Language Teachers through the Lens of EFL Students. *The Journal of Language Teaching and Learning*, 14(1), 1-23.

MLA

Mine Yıldız. "Language Tests and Assessment Practices of English Language Teachers through the Lens of EFL Students." The Journal of Language Teaching and Learning 14.1 (2024): 1-23.

Language Tests and Assessment Practices of English Language Teachers through the Lens of EFL Students

Mine Yıldız[1]

ABSTRACT

This qualitative study was conducted to reveal the common characteristics of language tests and the assessment practices of English language teachers from the perspective of language learners based on an analysis of the documents they produced. Namely, a series of cartoons created by 74 undergraduate students studying in the ELT department of a state university to reflect their views on and practices in English tests they had either taken or observed were examined and coded via thematic analysis. In line with the research questions, the findings were categorized under two main themes: (1) "the characteristics of language tests", and (2) "what language teachers do in assessment." The findings centered primarily on the negative aspects of language tests. In this sense, the evidence suggested that the language tests generally ignored the basic principles of assessment, such as reliability, validity, practicality, and washback effect. Similarly, negative attitudes of language teachers in the context of assessment were also portrayed in the cartoons. Given these findings, as well as the significance of language assessment literacy, particularly in practice, the need for language teachers to develop competence in language teaching and assessment is emphasized.

Language testing, as an inseparable part of language teaching, has been greatly influenced by many fields and theories related to language teaching and learning such as educational psychology and linguistics and four major approaches - (1) pre-scientific/essay-translation approach, (2) psychometric-structuralist approach, (3) psycholinguistic-sociolinguistic/integrative approach, and lastly, (4) communicative approach - to language testing have thus appeared. It has also advanced in three areas: (a) the theoretical perspective on language ability and the influence of test method and test taker characteristics on test performance, (b) the application of more sophisticated measurement tools, and (c) the development of

[1] Department of English Language Teaching, Atatürk University, Erzurum, Türkiye,
mine.yazici@atauni.edu.tr, +90 442 231 42 55

"communicative" language tests that incorporate principles of "communicative" language teaching (Bachman, 1991). Namely, language testing has evolved over time to become more humanistic while being less imposing, allowing for unbiased assessment of what learners do rather than solely reflecting on what they do/do not know (McNamara, 2000).

Assessment may serve a variety of purposes, including improving teaching and learning and enhancing school and student accountability as long as it is carried out effectively (Alderson & Wall, 1993; Bachman & Palmer, 1996; Brown, 2004). Thus, developing and conducting effective assessment procedures is invaluable for language teachers to diagnose the strengths and weaknesses in the language teaching process and thus design more effective teaching in the future. In this sense, Stiggins (1991) proposed the term "assessment literacy" as a requirement for teachers in all fields, as they are expected to produce high-quality tests and evaluate the assessment results critically. Teachers who are assessment-literate have the ability to design tests with a clear purpose, choose appropriate assessment techniques and tools in line with the goals of teaching and assessment, and thus carry out the assessment process effectively (Scarino, 2013). More specifically, the term *language assessment literacy* has gained importance in the context of language education. Language assessment literacy (LAL), which is necessary for not only language testing specialists and language teachers, but also educational policy developers and all test users, can be defined as "the acquisition of knowledge, skills, and principles of test construction; test interpretation, and use, test evaluation, and classroom-based assessments alongside the development of a critical stance about the functions of assessment within a larger educational context" (O'Loughlin, 2013, p. 363). A multi-facet concept, LAL involves the knowledge and competence of basic terms and principles of assessment, as well as using assessment types and tools appropriately to be able to design, develop, use, and evaluate language tests effectively (Coombe et al., 2020; Inbar-Lourie, 2008; Lam, 2015; Quilter & Gallini, 2000; Sultana, 2019). Therefore, assessment-literate language teachers are expected to be familiar with basic principles and concepts in testing and apply this knowledge "to classroom practices in general and specifically to issues related to assessing language" (Malone, 2013, p. 329). However, the existing studies concerning teachers' perceptions of and practices in LAL show that the ability of language teachers to design and conduct effective language tests is an ongoing problem. In this regard, although teachers may seem to have LAL in theory, their assessment practices often fail to reflect this competence (Babaii & Asadnia, 2019).

According to the National Education Association (1990), teachers should meet seven "standards for teacher competencies in educational assessment of students." Accordingly, teachers are required to be skilled in using and designing appropriate assessment methods and tools for their instructional decisions, as well as being adept at administering, scoring, and evaluating them appropriately. One of the best and most effective ways of enhancing language assessment literacy and thus satisfying these standards is to be educated about assessment. Therefore, nearly all teacher education programs in the world provide assessment courses in which students are intended to combine theoretical knowledge of assessment with practice. However, the effectiveness of such courses is still controversial, since both pre-service and in-service teachers suggest the content and procedures of these courses be revised to help them become more skilled in practice (Babaii & Asadnia, 2019; DeLuca & Klinger, 2010; DeLuca et al., 2015; Jeong, 2013; Jin, 2010; Lam, 2015). For instance, Jin (2010) found that classroom practice constituted just a small proportion of the instructional process, although the course instructors claimed to combine teacher lectures with student practice.

In terms of producing assessment tools, language tests designed and used by language teachers are required to demonstrate at least five basic principles, such as "practicality, authenticity, reliability, validity, and washback/backwash" (Brown, 2004; Hughes, 2003). As Brown (2004) suggests, a practical test should not be very costly; it should be administered within reasonable time restrictions; and it should include a specific and time-efficient scoring/evaluation technique. Moreover, through natural language, contextualized items, meaningful topics, and real-life tasks for learners, authenticity is achieved both for the test-users and test-

takers (Bachman, 2002; Brown, 2004). An effective language test should also be reliable in terms of raters, test administration, students, and the test, thus enabling consistency in measurement regarding the facets of a test condition. Furthermore, a reliable test is more likely to achieve another principle of an effective language test-- namely, validity -- since reliability is considered as a prerequisite of validity. In this respect, validity is simply the extent to which a test indeed measures what it aims to measure (Coombe et al., 2020; Hughes, 2003; Kaplan & Saccuzzo, 2004). Therefore, an appropriately-designed test should be well-balanced in terms of reliability and validity, since a loss in one will also lead to a loss in the other (Hughes, 2003).

A language teacher should also consider the effects of a test on all components of the teaching process, in addition to its design, administration, and evaluation; that is, an effective language test is also expected to have a positive washback effect for teachers and students (Bailey, 1996; Brown & Abeywickrama, 2010; Hughes, 2003; Messick, 1996). However, washback may sometimes be negative and thus hinder the learning process; for instance, failure on a test may discourage learners from studying, expecting further poor results (Ali & Hamid, 2020). Accordingly, all stages of language assessment (design, administration, scoring, and evaluation) should meet the criteria discussed here to achieve these principles interdependently, rather than independently. However, the existing studies investigating language tests and the LAL of language teachers emphasize that language tests prepared by even those teachers with theoretical LAL seem to lack these principles in practice (Babaii & Asadnia, 2019; DeLuca & Klinger, 2010; DeLuca et al., 2015; Jeong, 2013; Jin, 2010; Lam, 2015).

2. Literature Review

2.1. Language assessment in Türkiye

In order to provide students with better language learning outcomes, Türkiye has made a consistent effort to develop and conduct effective language teaching policies. With an updated curriculum revised in accordance with the general objectives of Turkish National Education, English lessons have been included in the educational program for all grades, from the 2nd to 12th. This new curriculum aims to improve the communicative competence of learners at earlier ages and emphasizes the need to remove the psychological barriers that hinder learning (Ministry of National Education [MNE], 2018). Therefore, it does not allow formal testing for learners at earlier ages as a means to prevent stress and anxiety that may lead learners to develop negative attitudes towards the language but emphasizes the use of alternative process-oriented testing techniques and self-assessment. However, formal assessment is not proposed to be totally put aside; rather, students' achievement is required to be assessed through different techniques such as written and oral exams, quizzes, homework assignments and projects.

Language teaching in Türkiye, where education is still exam-oriented, mostly depends on formal testing to make final determinations about language learners' progress (Hatipoğlu, 2010). Although both the curriculum and the coursebooks used in English courses emphasize integrated teaching of the four language skills, in practice, language tests are most often based on the measurement of receptive skills and linguistic knowledge of learners through out-of-date written examinations consisting of wh- questions, fill-in-the-blanks items, and multiple-choice questions (Haznedar, 2012; Kirkgöz et al., 2017; Köksal, 2004; Ölmezer Öztürk & Aydin, 2018; Sarıçoban, 2011). More interestingly, the proficiency of learners in productive skills is assessed through tests consisting of restricted-response items. For instance, in his study investigating English tests designed or used by English teachers in Türkiye, Sariçoban (2011) pointed out that "instead of marking speaking, the teacher gives marks according to the worksheets, homework and participation in the lesson by giving plus and minus signs" (p. 405).

It has been recommended that the traditional paper-pen tests and summative assessments on which language assessment heavily depends be replaced or supplemented by alternative assessment tools and formative assessment in order to effectively evaluate the language proficiency of learners in all skills (Kirkgöz, 2007; Öz, 2014). In this respect, formative assessment and other alternative assessment tools (i.e.,

portfolios, projects, etc.) are believed to be more effective particularly for young learners (McKay, 2006), since they are more compatible with the principles of communicative competence proposed to be gained in language classes. A gap between teaching and assessment is no longer acceptable since an effective assessment process is expected to be parallel with the language teaching process; that is, teachers are required to assess just what they teach and in the same way that they teach (Köksal, 2004). Therefore, language teachers must also be well-trained in assessment in addition to language teaching. However, studies point out that teachers generally attribute their insufficiency in language assessment to the lack of practical assessment knowledge they have received in their pre-service education (Hatipoğlu, 2015; Köksal, 2004; Mede & Atay, 2017; Ölmezer Öztürk & Aydın, 2019; Öz & Atay, 2017)

Since 2007, all institutions of teacher training in Türkiye have been obligated to follow the same teacher training program, in which only the elective course offerings may vary. According to the Teacher Education Program that was applied up to 2018, all pre-service English teachers were required to take the course titled Assessment and Evaluation in Foreign Language Education in the seventh term of their teacher education programs (Council of Higher Education [CHE], 2007). With certain revisions to the program in 2018, the course was moved to the final term of the program with a new title, "Designing Tests in English Language Teaching," but with the same purpose (CHE, 2018). This course basically aims to help pre-service language teachers to become competent in language assessment in order to "demonstrate to stakeholders that the intended use of their assessments is justified" (Bachman & Palmer, 2012, p. 2) when they begin teaching English at any level. Although the program provides a basic description of the components to be involved in the course, the instructor conducting the course may influence the process and outcomes of the course (Jeong, 2013; Ölmezer Öztürk & Aydın, 2019).

Various studies investigating the effectiveness of assessment courses (which may be offered under various course titles) from the perspectives of in-service and pre-service teachers (Hatipoğlu, 2010; Öz & Atay, 2017) demonstrate that such courses help language teachers develop awareness and more profound understanding of the complex concepts and theoretical knowledge in assessment, but that more practical opportunities balanced with theory would be preferred. For instance, Öz and Atay (2017), in their study aiming at revealing Turkish EFL instructors' perceptions of and practices in language assessment, and thus reflecting the relationship between their perception and own experience, emphasized the similar conclusion that "there is a mismatch between Turkish EFL instructors' in-class assessment literacy and its reflection in practice" (p. 39).

Studies that have examined the characteristics of language tests and assessment procedures applied in the Turkish context reveal that language assessment principles are largely ignored (Kirkgöz et al., 2017; Köksal, 2004; Sarıçoban, 2011). In some instances, language teachers expressed the belief that they had gained language assessment literacy in the related courses in their teacher training programs; however, the studies emphasized a gap between their perceptions and practices (Hatipoğlu, 2010; Öz & Atay, 2017).

Overall, the studies found in the literature generally focus on language assessment process from the perspective of in-service and/or pre-service language teachers as a means to reveal their competence in language assessment (Köksal, 2004; Mede & Atay, 2017; Ölmezer Öztürk & Aydin, 2018; Ölmezer Öztürk & Aydın, 2019; Sarıçoban, 2011). However, there are other stakeholders in this regard, namely students themselves, who are the test-takers (Ricci et al., 2018). Therefore, their views about and experiences in language assessment are also valuable as a means to improve the language assessment process and outcomes. In this regard, it may be difficult and frustrating for test users to express the weaknesses and errors in language assessment; yet these may be more clearly and critically seen from the counter side; e.g., by the test-takers themselves. With this in mind, in the course titled "Assessment and Evaluation in Foreign Language Education" conducted by the researcher in the English Language Teaching (ELT) department of a state university in Türkiye, the students, though being teacher-candidates, complained about the tests they had been exposed in the past after they were instructed about topics such as the basic terms; principles

of testing, assessment, and evaluation; and key points to be considered in the design, development, administration, and evaluation of effective language tests. In this sense, they report that "the principles sound meaningful, but what we have experienced and are even still experiencing is totally different." In this course, students are asked to design, administer, and evaluate a language test based on five principles of testing, assessment, and evaluation, such as practicality, authenticity, reliability, validity, and washback/backwash effect. Thus, they report gaining considerable experience in language assessment. On the other hand, the more they learn about language assessment, the more they criticize the current and previous testing processes they have experienced. Moreover, several of the students in the study requested that the researcher share the knowledge of language assessment with her colleagues, in addition to in-service language teachers whose assessment they thought still lacked these principles. As a result of such discussions, the students taking this course in the eighth and final term of their teacher education program in the 2020 academic year were asked to reflect on the language assessment practices and tests they had observed or been exposed to by designing cartoons on the website "www.makingbeliefscomix.com".

The present study was conducted to reveal the common characteristics of language tests and assessment practices of language teachers from this perspective. In line with this aim, students' experiences with and perceptions about language tests were reflected in cartoons and were analyzed in order to answer the following research questions:

1. How do EFL students portray the English language tests that they have observed or taken?
2. How do EFL students react on their English language testing experience?

3. Method

3.1 Research design

In qualitative research, everything that pertains to a given topic can be potential data (Dörnyei, 2007). As a part of this qualitative study, cartoons created by the participants were considered valuable documents, with the advantage of revealing their thoughts and experiences relating to language assessment in their own language and words (Cresswell, 2012). Unlike traditional methods relying on verbal and written methods, such visual methods are also suggested to "enhance the richness of data by discovering additional layers of meaning, adding validity and depth, and creating knowledge" and "result in increased trustworthiness of the findings" (Glaw et al., 2017, p. 1). Taking into account all these strengths of visual document analysis, the researcher used cartoons drawn by participants to express their experience with and views on the English tests to uncover meaning, acquire knowledge, and discover insights pertinent to the research problem (Merriam, 1988). In addition, document analysis allowed the researcher to gather the perspectives and experiences of more individuals than would be possible through interviews, as well as more specific and trustworthy data than would be gained through questionnaires.

3.2. Participants

The participants in this study consisted of 74 pre-service English language teachers (19 males, 55 females) at a state university in Türkiye. Their ages ranged between 22 and 31 years. With the exception of three seniors, all of the students were enrolled in the final term of their undergraduate education. Therefore, during the preceding seven terms, they had taken numerous theoretical and practical courses in language teaching. Among these was the course on testing and assessment, which included basic theoretical concepts with regard to testing and assessment. In this context, they were not only students who were still required to pass classroom tests, as well as large-scale tests; they were also pre-service teachers who had the opportunity to observe in-service language teachers and practice their language teaching (although in a very limited manner due to the pandemic) in their teaching practicum courses. Since the participants had different educational backgrounds, from colleges to state schools in different cities of the country, the

participants' reflections provide a good representation of language assessment practices generally conducted in Türkiye.

Table 1

Demographic information about participants

Category	Subcategory	Frequency (N)	Percent (%)
Gender	Male	19	25.7%
	Female	55	74.3%
Grade	Seniors	3	4.1%
	Juniors	71	95.9%
Age	Min.	Max.	Mean
	22	31	23.4

3.3. Data collection

The data were collected through the analysis of a series of cartoons the students designed on the website "www.makingbeliefscomix.com" to reflect their perceptions on and experience with the English tests they have taken or observed in English classes during their own learning processes and practicum. This website, created by Bill Zimmerman, enables users to express their ideas or stories in comics in a number of languages. In the 11ᵗʰ week of the course, when the teaching of theoretical knowledge in language assessment had been completed, the students were first asked to orally discuss common language assessment practices in Türkiye that they had observed or encountered. Then, they were assigned to reflect on their experiences or observations in language assessment through cartoons using the website above. They were also informed about how to use the website. They completed their assignment in one week and then shared it on Moodle.

3. 4. Data analysis

Each cartoon created by the participants within the scope of an assignment was first analyzed by the researcher as an instructor of their course in terms of the requirements of the assignment. Students were given written feedback on their cartoons, which were orally discussed in the following lesson with the students and member checking was thus endorsed. Four of the submissions were excluded because they had different content (i.e., reflecting on the effectiveness of the course "Assessment and Evaluation in Foreign Language Education"). All of the other assignments were selected to be included in data analysis and thus downloaded. The researcher analyzed the students' assignments (described as the documents for the study) with the help of N-Vivo 12 Pro based on a thematic analysis procedure according to commonalities (Gibson & Brown, 2009). In this process, the students' cartoons were first examined for their commonalities such as those regarding the principles that an effective language test should achieve and others reflecting what they had experienced in English assessment, and empirical codes were obtained accordingly. The relationships between codes were then examined and included in a code family. The researcher analyzed the data four weeks later in order not "to become 'stuck' or 'frozen' on some intermediate construction [interpretation]" (Guba and Lincoln, 1989, p 238) and to enable progressive subjectivity; as a result, nearly identical codes were obtained. For the credibility of findings, the codes were also examined and checked by another researcher who had given lectures on language assessment at undergraduate and graduate levels. Furthermore, sample cartoons were given for each code to ensure that

findings are not colored by the biases of the researcher but accurately portray the responses of the participants.

The code family is illustrated in Figure 1. Accordingly, two main categories were obtained. While the former reflects what they think about language tests, the latter is about what they have experienced in language assessment.
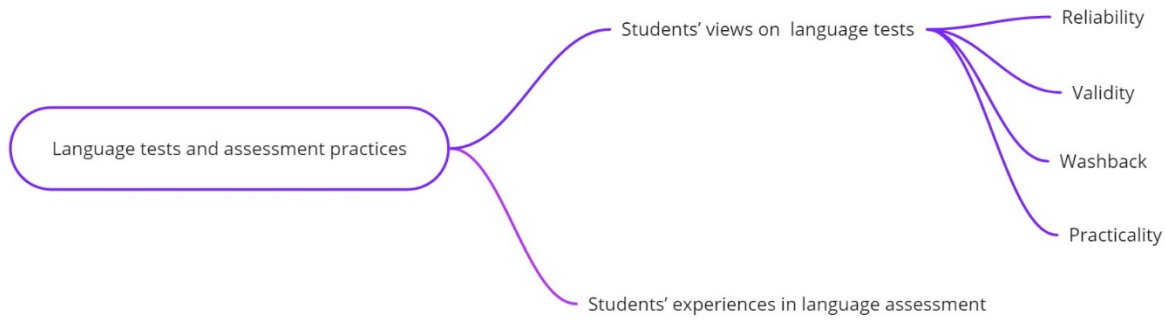

Figure 1. The code family of the study

4. Findings and Discussion

In this study, the students prepared cartoons on a website to reflect their views on language tests and the assessment practices of language teachers that they had experienced in their own language learning processes or they had observed in their practicum. The findings were obtained on the basis of a detailed analysis of the cartoons and are presented here with examples, then discussed in light of the research questions. Although their cartoons included some structural or mechanical errors, these were not corrected by the researcher in order not to interfere with their originality.

4.1. Students' views on language tests

When the students' cartoons were analyzed, it was found that they mainly focused on four of the basic principles that must be demonstrated in an effective language test: reliability, validity, washback, and practicality. Almost all of the cartoons reflected the weaknesses of the tests in terms of these principles; just a few involved positive reflections about tests conducted by language teachers.

4.1.1. Reliability

One of the basic principles of an effective language test is reliability. A reliable test is one that provides consistent results when it is administered to different students under similar circumstances or to the same students on different occasions (Brown, 2004; Frankel et al., 2012). Furthermore, the scoring and evaluation process should also be consistent in order to obtain reliable results (Bachman & Palmer, 1996; Brindley, 2001; Frankel et al., 2012; Weigle, 2002). On the other hand, in this study, the unreliability of teachers' scoring was largely emphasized by the students. In this regard, they generally constructed at least one caricature showing language teachers' unfair and unreliable scoring and evaluation practices. For example, they indicated that their teachers generally gave the same students fixed scores, either consistently high whether they did well or not; or consistently low, even if they performed better. Moreover, they expressed that their teachers had favorite students who did not need to study for their exams, because they always got high scores. This was illustrated in the following examples:
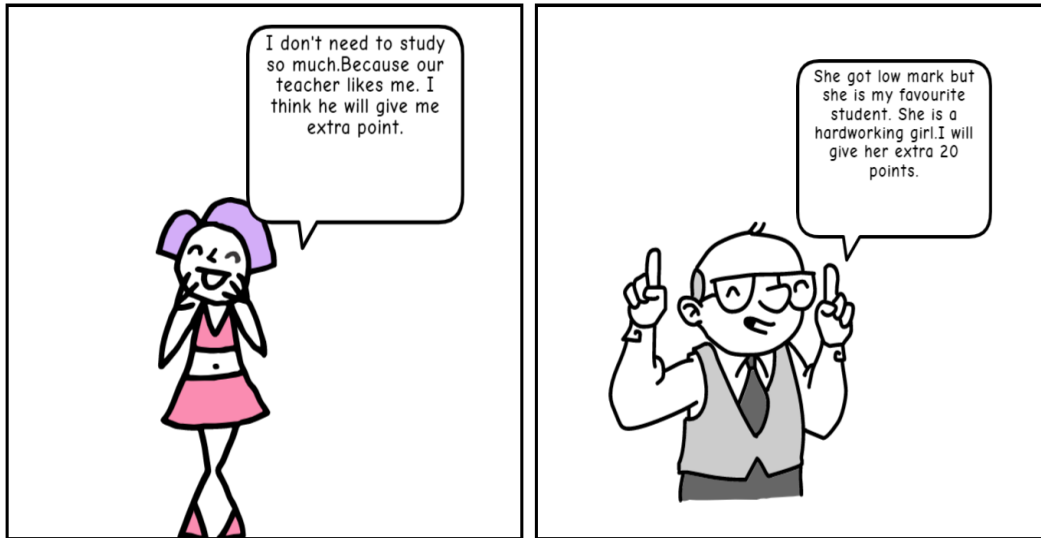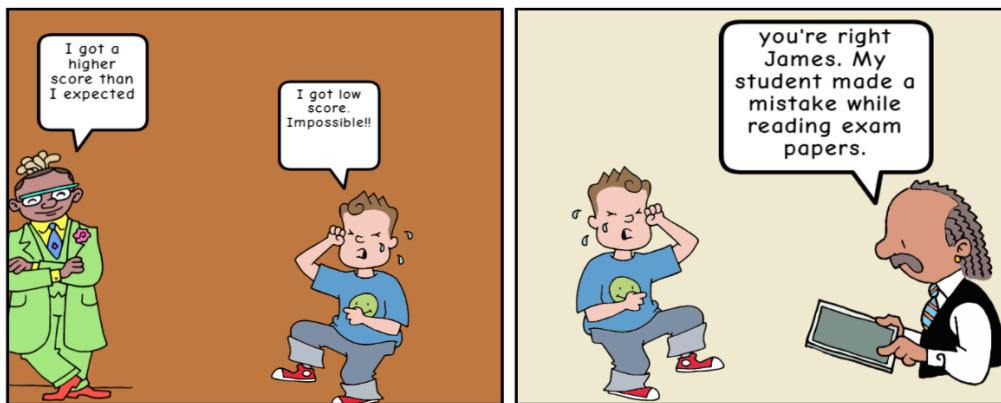
(S11).

Furthermore, the students also illustrated that a teacher might give different scores for the same response or get someone else (e.g., their assistants or students) to score the results, as seen in the following examples:

T: The first question for this paper is 10 points.
T: The first question for this paper is 30 points (S38).

(S41)

Another issue addressed by the students that led them to describe their teachers' scoring and evaluation as unreliable was that they were not permitted to see their papers after evaluation. They claimed that their teachers did not allow them to see how their papers were scored or what their errors were.

(S22)

Moreover, if they asked to see their papers, they might experience harsh reactions from their teachers. Some of the cartoons illustrated that their teacher might even warn them about taking away points. Thus, besides missing the opportunity to learn from their errors, they were left with suspicion about the evaluation of their papers. This case was reflected in the following comic:
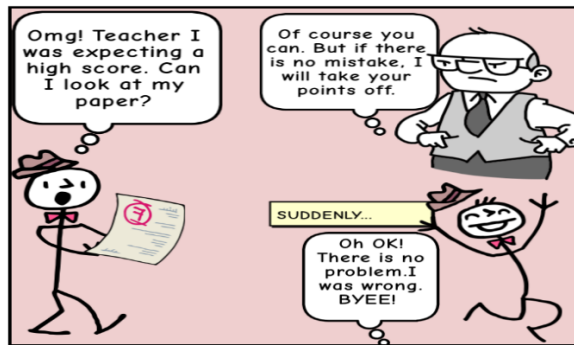

(S 13)

On the other hand, some of the cartoons pointed out the opposite case. For instance, they reflected that their teachers allowed students to see their papers and thus notice any errors or mistakes made in scoring, as in the following example:
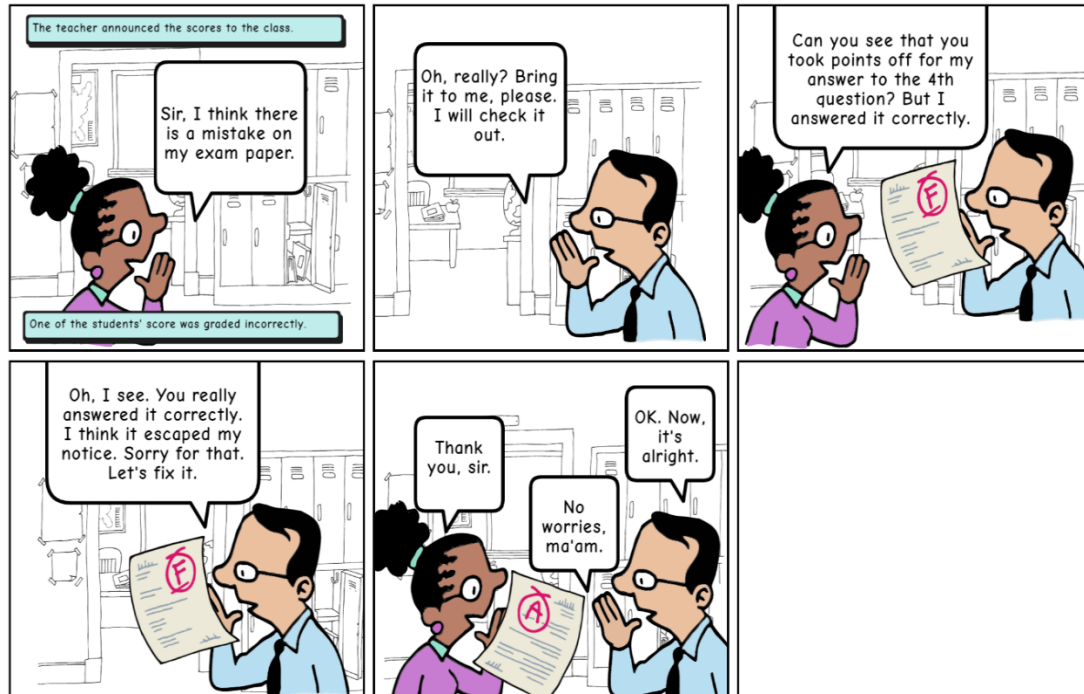
(S49)

Although the students primarily touched on problems with rater reliability, they also emphasized issues regarding test and student reliability. For instance, they claimed that the use of multiple-choice questions on language tests might yield inconsistent results, since the students could easily cheat, as exemplified in the following:
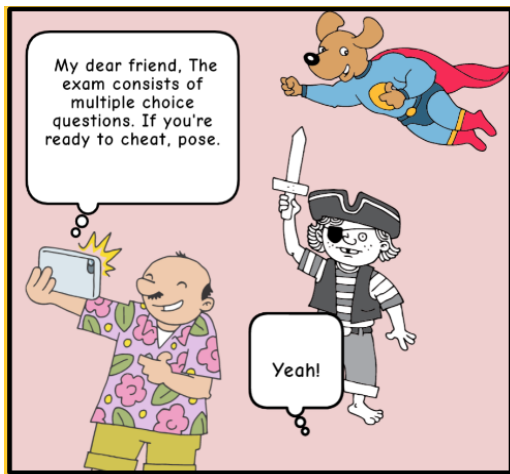
(S18)

The students also expressed that their teachers ignored factors regarding the test-takers that might greatly affect the results of the tests (Brown, 2004). For instance, even if they got wet in the rain on the way to school or caught a terrible flu, they were required to complete a test, as demonstrated by S14:
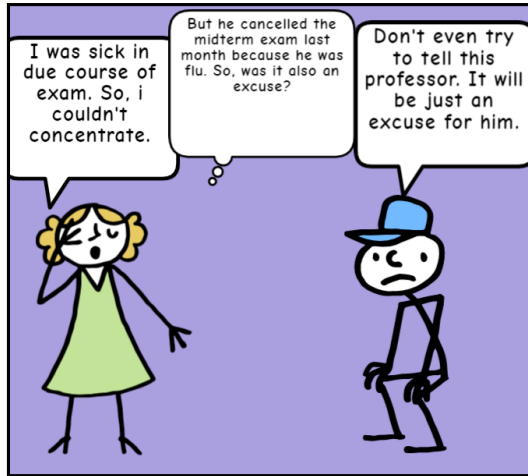
(S14).

As the findings reveal, reliability issues regarding the test itself, raters, administration, and students are seen by students as being ignored by teachers in classroom language tests by the teachers. Furthermore, the biases, thoughts, and beliefs of the teachers may also greatly shape the results (Brown, 2004; Hughes, 2003; Zhang, 2016). For example, a student whom the teacher always admires in class may be perceived as having no need to study for the exam since s/he will undoubtedly receive the best score no matter what his/her responses may be. Additionally, the students asserted that language teachers do not favor allowing students see their papers or check their results when the students think that the results might be wrong. Likewise, in addition to rater reliability, administration and student reliability can also be ignored by language teachers. In this respect, while teachers may prefer the multiple-choice approach, perhaps in order to provide rater-reliability, they may overlook the fact that multiple-choice questions may conflict with administration reliability, since students can easily cheat on such an exam. These results are also supported by Haznedar (2012), Köksal (2004), and Sarıçoban (2011), who found that language teachers frequently prefer multiple-choice items to assess language performances of learners, which may result in problems for the validity, and thus the reliability of the test. However, Kirkgöz et al. (2017) presented partially contradictory results showing that tests using matching items are more common than multiple-choice questions.

4.1.2. Validity

As a prerequisite for reliability, validity simply refers to the extent to which a test measures what it is supposed to measure, and nothing else (Brown, 2004; Ching, 2019; Gipps, 1994; Heaton, 1990; Hughes, 2003; Kaplan & Saccuzzo, 2004). Although this is dealt with according to the evidence on which it depends, such as content-related, criterion-related, construct-related, and consequence-related validity, the students in this case mainly focused on issues of content-related validity. In this sense, an effective test should be a good representation of the topic taught in the course in order to have content validity (Weir, 2005). However, the students claimed that their tests did not cover only what they learned; in other words, it is possible to encounter topics on a test that had not been covered in the course content, as exemplified in the following comics:
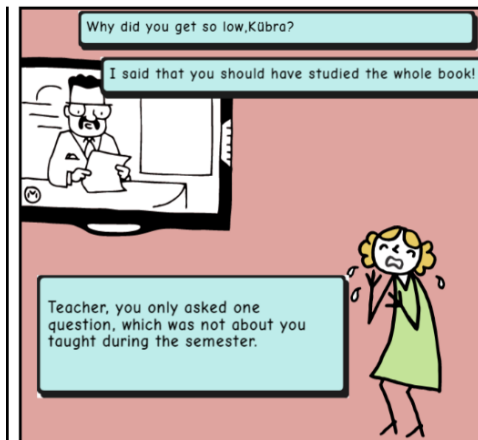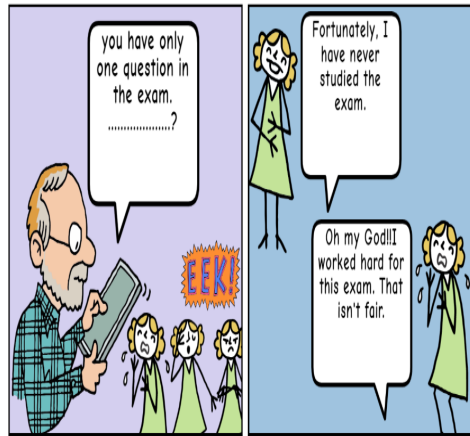
(S21) (S12)

Furthermore, the cartoons produced by the students showed a perception that tests do not measure proficiency in the target language in terms of the communicative and linguistic competence of learners. Moreover, they reflected that their tests tended to lack construct validity, which refers to whether learning language theories or constructs involved in language teaching also appear in language assessment (Bachman, 2002; Brown, 2004). On the contrary, tests generally consisted of multiple-choice questions measuring their reading skills, grammar, and vocabulary knowledge. As such, they needed to memorize the rules or words they learned, rather than learning how to use them. These issues are illustrated in the following examples:
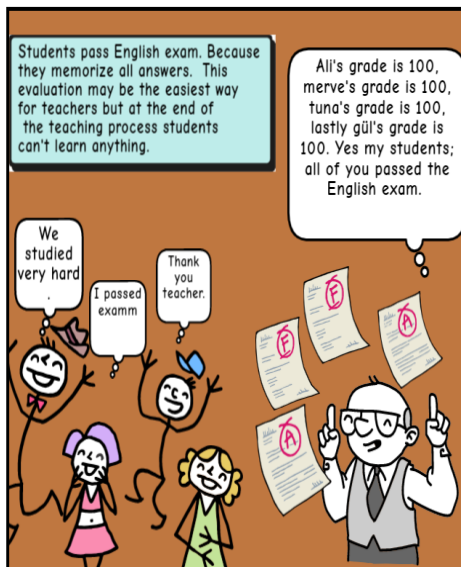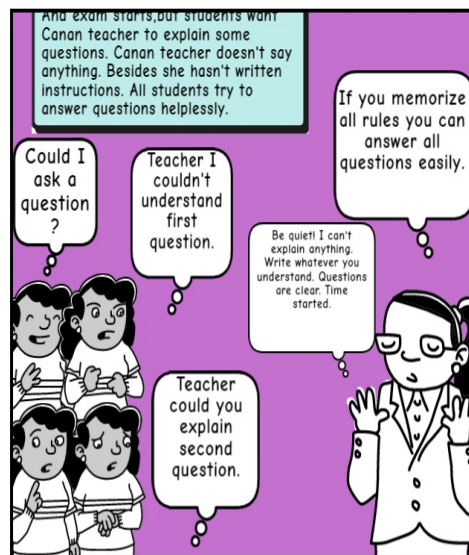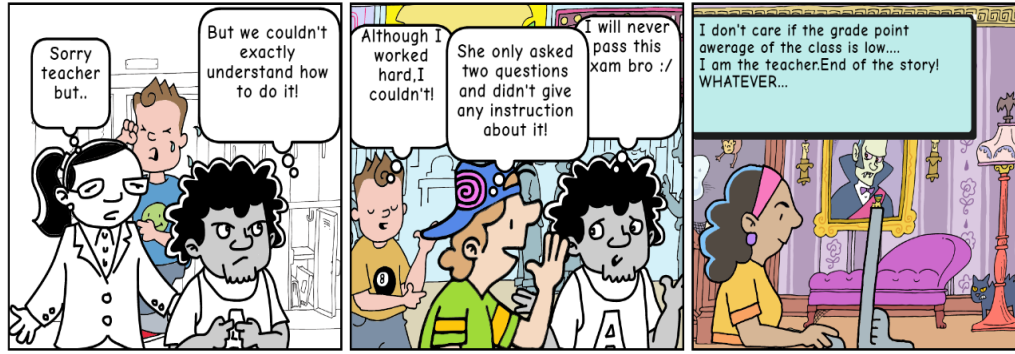

(S8) (S64)

The students' statements in the cartoons also pointed out that consequence validity, which pertains to the consequences of the tests for test-users, test-takers, the teaching process, and all other stakeholders in education (Brown, 2004; Gipps, 1994), was ignored by language teachers; this concern is illustrated by one of the students as follows.

(S58).

As the findings show, the cartoons created by the students illustrated that language tests prepared by English language teachers generally suffer from validity, and content validity in particular. For instance, they complained about mismatches between what they learned in class and what they were asked on their exams. Furthermore, their proficiency in the target language was assessed only through restricted-response items designed for the purpose of measuring receptive skills: particularly the reading, and grammar and vocabulary knowledge of learners (Kirkgöz, 2007). Sarıçoban (2011) similarly pointed out that the language tests evaluated in his study primarily covered reading, vocabulary, and grammar skills, although writing and listening were also included in the coursebook. As such, his study supports the findings of this study suggesting that language tests generally focus on recognition, rather than the production of learners, and thus ignore productive skills such as writing and speaking "by having them write and their speaking skills by having them speak" (Köksal, 2004, p. 4). Aside from content and construct validity, consequence validity was also viewed as being pushed aside by the teacher. However, students think that their language teachers should notice the fact that language tests may have a great influence on their lives from a variety of perspectives, such as eligibility for scholarships or higher education (Hatipoğlu, 2010).

4.1.3. Washback

A particular aspect of consequence-related validity, washback/backwash effect, which refers to the influence of assessment on each component and stakeholder of language teaching (Bachman & Palmer, 1996; Bailey, 1996; Brown, 2004; Brown & Abeywickrama, 2010; Hughes, 2003; Messick, 1996), was also emphasized by the students in the study. According to Alderson and Wall (1993); Ali and Hamid (2020), and (Hughes, 2003), washback may not always be positive, in contrast to what is expected. For instance, as reported by the students, test results may demotivate them and even lead them to develop negative attitudes towards the language. For example, when they were exposed to such a situation wherein they had studied harder for an exam or performed similarly on the exam to their friends and got lower grades, they might give up studying or even listening to their teachers. The following statements in their cartoons illustrate these views:
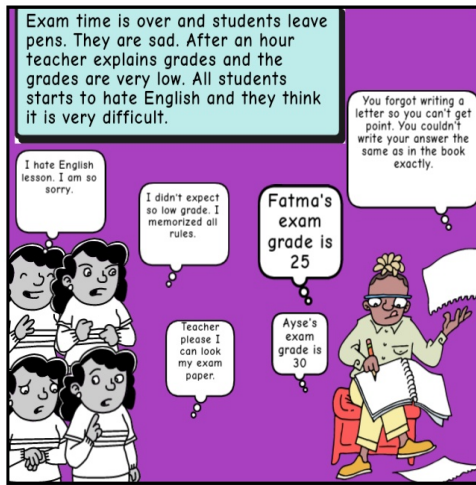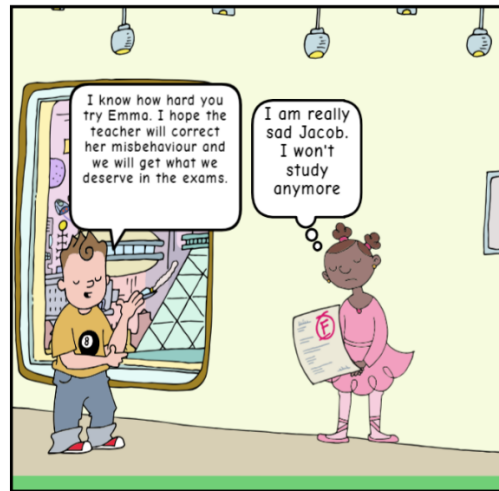
(S64)   (S27).

The students' reflections in their cartoons demonstrate a belief that language tests may yield a negative washback effect for learners. In other words, language test results or the teacher's attitudes during an assessment may demotivate the students and thus hinder their learning. As they expressed, when they did not perform well even after studying hard, they began to consider that they could not succeed and that they would not learn English anymore. In support of this view, Alderson and Wall (1993) state that a test influences what and how a teacher teaches as much as what and how a student learns, in addition to affecting attitudes towards the content, method, and learning. These findings reveal that negative results on tests lead students to be disinclined to go on learning a language.
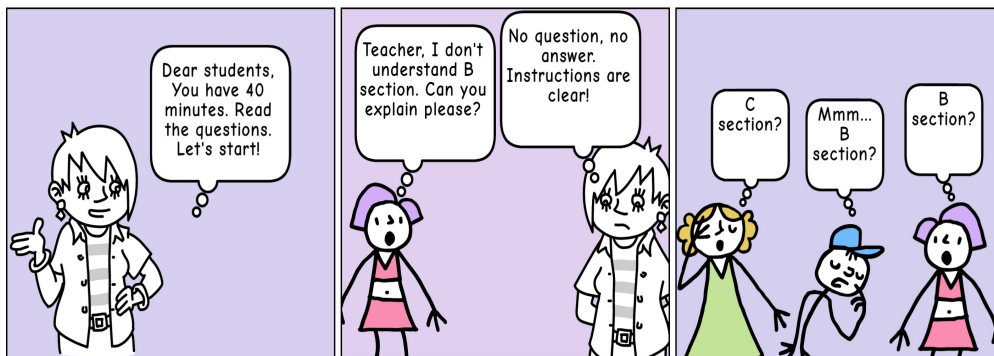
4.1.4. Practicality

Practicality is defined as "the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities" (Bachman & Palmer, 1996, p. 36). With this in mind, a practical language test should not be too expensive; it should not consume too much time to design, complete, and evaluate; it should not be too difficult to administer; and it should have unambiguous scoring and evaluation procedures (Brown, 2004). However, in this study, the students revealed that they had encountered many language tests that did not include any instructions about time or scoring. In their cartoons, they illustrated that their teachers simply started the testing procedure and then said "time is up" when it was finished, rather than informing students how much time was allocated to complete the test. Furthermore, their language tests did not include any information about the questions would be evaluated and scored. The story of S1, for instance, exemplifies the students' complaints about the impracticality of language tests in terms of time and scoring/evaluation procedures:
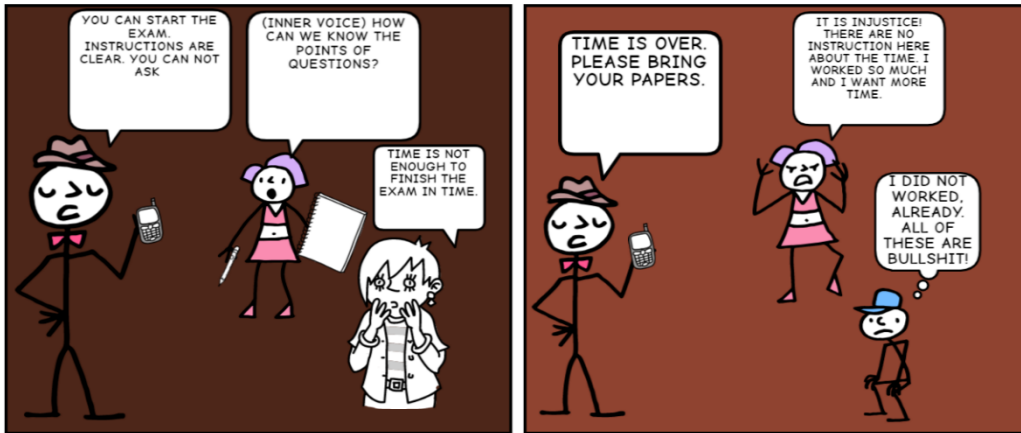
[S1]
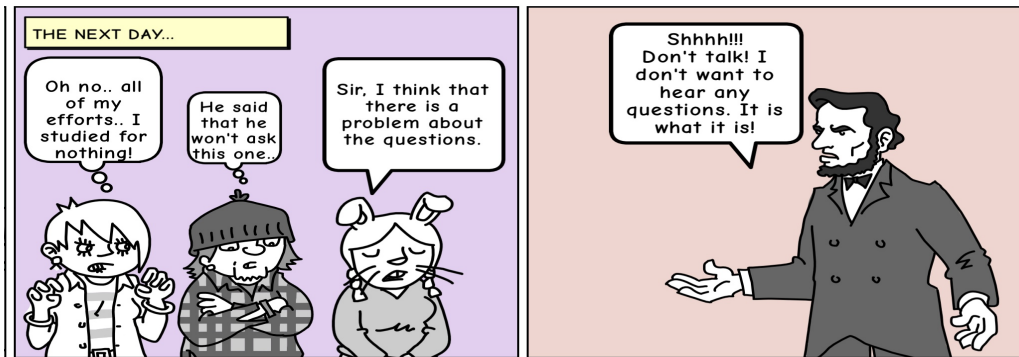
(S21)

(S44)

In this sense, the students expressed that language tests do not adhere to the principle of practicality because the time allocated to a test may be insufficient to complete it. Furthermore, language teachers may fail to include instructions indicating the time allotted for taking the test and explaining the scoring procedure, including the number of points assigned to each question. This finding is similar to Köksal (2004), who also pointed out that the language tests analyzed in his study generally did not involve any information about "how much time was allocated; or what points the test-takers would get for each correct response" (p. 5). According to (Schoonen, 2011), the inclusion of clear and explicit instructions on a language test in terms of

how students are required to respond to the questions, how much time they have to complete the test and how their responses will be scored not only impacts its practicality, but also enhances its reliability and validity.

4.2. *Students' language assessment experiences*

The cartoons produced by the students' cartoons reflected their language assessment experiences and thus gave information about their language teachers' behaviors and attitudes during exams. In this regard, the students mainly criticized their teachers' negative and demotivating behaviors with cartoons illustrating that language teachers did not allow students to ask questions about the tests and neglected to provide clear instructions regarding time allocation, scoring procedures, and response format. Moreover, they pointed out that teachers in some instances responded harshly to questions or objections from students. These types of teacher behaviors undoubtedly affect students' testing performance and results, as well as their learning process and even attitudes toward the language. Moreover, when teachers behave harshly during exams, students become more stressed and anxious about their test performance. The following statements illustrate these concerns:
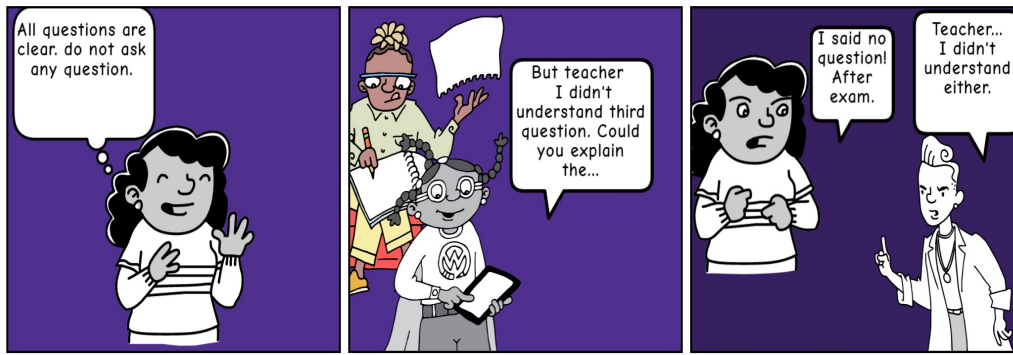
(S47)

Furthermore, through their cartoons, some of the students implied that teachers might take satisfaction from students having difficulty giving responses to their test questions, believing this would motivate them to study harder. In addition, they expressed that their teachers were not concerned about the test results or their role as indicators for the teaching process.

From another perspective, the students also criticized some of the assessment practices of their language teachers, such as conducting unannounced exams or asking different questions for different groups on the same test, as presented in the following examples:
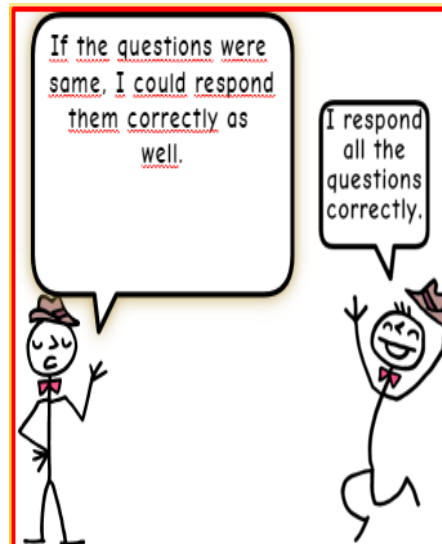
(S70)   (S4)

On the other hand, they also pointed out some positive implementations from language teachers in terms of assessment. For instance, in some examples, language teachers evaluated their students based on in-class performance or alternative assessment tools, rather than relying solely on formal tests. Moreover, they implied that novice teachers or those who found their language assessment literacy to be insufficient might get assistance from experienced or more assessment-literate teachers. In addition, they indicated that some teachers shaped their teaching process and future language tests on the basis of students' test results or feedback. All these three situations were exemplified through the following comics:
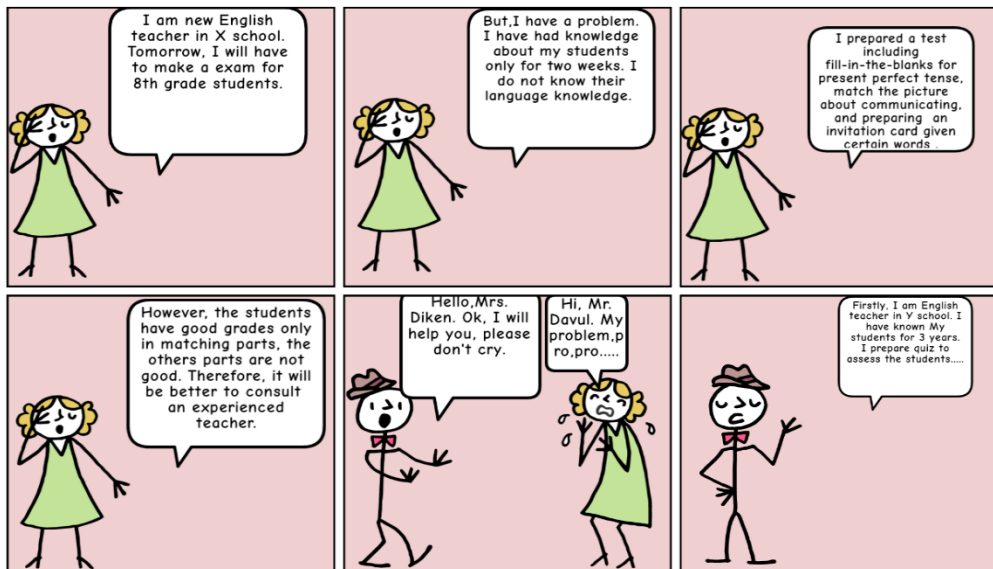
(S59)
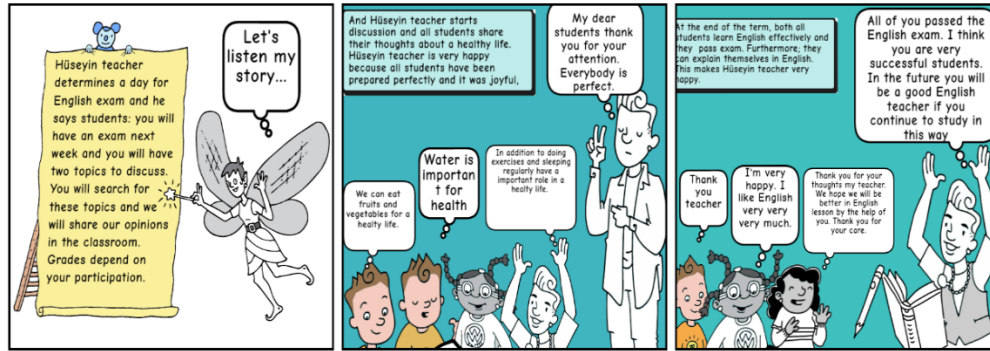
(S24)

The students' stories also reflected language teachers' practices in assessment, indicating that language teachers tended to ignore the principles of an effective language test in practice, even if they knew these principles in theory. In this sense, the students felt that teachers believe that the more challenging an exam is, the harder students will study. In addition, teachers' efforts to discipline students during the exams were seen as increasing their anxiety and stress, adversely affecting their performance. However, it cannot be said that all language teachers' assessment practices were erroneous; on the contrary, some language teachers were said to struggle to design better language tests discussing with their colleagues. They also considered the tests' washback effect on their teaching process and on motivating their students about their future learning.

In support of the previous studies, the importance of "language assessment literacy" for language teachers was also detected in this study (Coombe et al., 2020; Inbar-Lourie, 2008; Lam, 2015; Quilter & Gallini, 2000; Sultana, 2019). Language assessment literacy, which is suggested as one of the competencies that a teacher should have, enables language teachers to design, conduct, and evaluate language tests effectively (NEA, 1990). However, several studies point out that while teachers are equipped with LAL in theoretical terms through their coursework, their assessment practices in the classroom do not reflect their knowledge about assessment and evaluation (Babaii & Asadnia, 2019; DeLuca & Klinger, 2010; DeLuca et al., 2015; Jeong, 2013; Jin, 2010; Lam, 2015). Similarly, the findings of this study reveal that students see language teachers as having difficulties with issues such as designing, conducting, and evaluating effective language tests; assessing their learners' performance and proficiency in the language; and shaping their further teaching process based on the results of their assessments.

This study also presents evidence that teachers' assessment practices affect not only the students' performance on tests, but also impact their future studies and even their attitudes toward the course. As such, language teachers' assessment practices may have a more comprehensive and sometimes long-lasting effect, since these may influence students' attitudes toward the language throughout their lives.

As much as being a significant source of anxiety and stress, language tests also stand as playing an essential role in motivating students. As suggested by the participants, the tests' lack of ability to measure and evaluate their knowledge and performance discourages students from studying and may even lead them to think they cannot succeed. Therefore, it is evident that there is a strong need for language teachers to have and apply the LAL in all their assessment procedures, not only in terms of assessment and evaluation but in language teaching overall, as well.

5. Conclusion

This study was conducted to examine students' views on language tests and language assessment experiences. Accordingly, the students taking the course entitled "Assessment and Evaluation in Foreign Language Education" in the ELT department of a state university in Türkiye were asked to design cartoons

to reflect the language assessment practices they had experienced or observed. The cartoons were created on a website assigned by the instructor. The researcher, who also taught the course, analyzed their cartoons based on the document analysis technique to reveal the common themes they revealed.

According to the findings of the study, language tests designed and/or applied in Türkiye are generally believed by students to be deficient in the basic principles of assessment. Since teachers reflect their beliefs, thoughts, and biases in their evaluation practices, reliability issues may emerge. Likewise, their over-dependence on certain assessment techniques, such as multiple-choice questions, also decreases the reliability of their tests. In addition, it is also emphasized that language tests ignore the principle of validity; namely, there is often a mismatch between what is taught in class and what is included in an exam. For instance, although learners are typically asked to perform their speaking and writing skills during lessons, exams offer no opportunities for written or oral production.

Similarly, it was observed that there is often little to no clarification of the duration and scoring procedures for language tests, which diminishes the practicality of the test. Likewise, the principle of washback effect is also claimed to be ignored by language teachers. As such, language tests generally result in negative washback for learners in terms of applying themselves to the learning process and their attitudes toward language learning. In this sense, when exposed to undesired situations during or after a test, they lose motivation for learning the language.

As for the students' language assessment experiences, the findings similarly focus on language teachers' negative practices in assessment. However, certain positive approaches of language teachers, such as consulting with colleagues about a test they have designed and considering and evaluating test results to motivate the learners and shape their teaching, were also pointed out.

In line with these findings, it can be concluded that effective language assessment and evaluation is vital to advance language teaching (Sultana, 2019), since teachers' language assessment practices can shape the learning process, as well as learners' attitudes towards the language. In other words, the characteristics of language tests and the assessment practices of language teachers have a significant impact on language teaching as a whole, in addition to their effects on assessment and evaluation. However, as pointed out in this study, students believe that language teachers struggle with issues such as designing, conducting, and evaluating effective language tests; assessing their learners' performance and proficiency in the language; and shaping their subsequent teaching process based on the results of their assessments. Therefore, this study provides valuable insights for both pre-service and in-service language teachers about what they should and should not do while planning, using, and administering a language test. In addition, language teachers, from the findings of this study, can see the reflections and impacts of what they do in language assessment on their students, and thus improve what they do wrong while sustaining what they do best.

As a result, this study emphasizes the importance of English language teachers being well-educated on assessment and developing adequate language assessment literacy skills to design and use effective language tests for measuring the actual performance of learners in the language, as well as evaluating their own teaching. The language assessment course provided in teacher education programs has tremendous importance in the education of language teachers; however, this course should not be confined to theory alone and pre-service language teachers should be provided with more opportunities to observe, practice, and evaluate language assessment. Furthermore, assessment education should always be a part of the professional development of in-service language teachers. As done in this study, language teachers may also discuss their assessment practice with their test-takers after assessment and evaluate their tests and assessment skills from the other perspective, and thus notice what they should or should not do in their future assessment and teaching.

Although student cartoons have provided valuable insights for this study by revealing the common characteristics of language tests and the assessment practices of language teachers, they reflected only the test-takers' perspectives and experiences. To gain further insight, language tests conducted by language

teachers may be objectively investigated to see whether they consider the important language assessment principles. Furthermore, evaluating learners' and teachers' experiences and views about assessment may also shed light on the status of language assessment in Türkiye in detail.

Funding

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied linguistics, 14*(2), 115-129. https://files.eric.ed.gov/fulltext/ED345513.pdf.

Ali, M. M., & Hamid, M. O. (2020). Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh? *Language Assessment Quarterly, 17*(2), 129-146. https://doi.org/10.1080/15434303.2020.1717495

Babaii, E., & Asadnia, F. (2019). A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice. *Reflective Practice, 20*(6), 745-760. https://doi.org/10.1080/14623943.2019.1688779

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4), 671. doi:10.2307/3587082

Bachman, L. F. (2002). Alternative interpretations of alternative assessments: Some validity issues in educational performance assessments. *Educational measurement: Issues and practice, 21*(3), 5-18. https://doi.org/10.1111/j.1745-3992.2002.tb00095.x

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Bachman, L., & Palmer, A. (2012). *Language assessment in practice*. Oxford University Press.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(3), 257-279. https://doi.org/10.1177/026553229601300303

Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (Vol. 11, pp. 137-143). Cambridge University Press.

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. https://doi.org/10.1080/0969594042000304609

Brown, H. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. Pearson Education White Plains.

Ching, G. (2019). *Teaching English: A practical guide for language teachers*. Canadian Scholars.

CHE (2007). Education faculties teacher training programmes. Retrieved from https://www.yok.gov.tr/Documents/Yayinlar/Yayinlarimiz/egitim-fakultesi-ogretmen-yetistirme-lisans-programlari.pdf

CHE (2018). English Language Teaching undergraduate programme. Retrieved from https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Yeni-Ogretmen-Yetistirme-Lisans-Programlari/Ingilizce_Ogretmenligi_Lisans_Programi.pdf

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn? *Language Testing in Asia, 10*(1). https://doi.org/10.1186/s40468-020-00101-6

Cresswell, J. W. (2012). *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*(4), 419-438. https://doi.org/10.1080/0969594x.2010.516643

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2015). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability, 28*(3), 251-272. https://doi.org/10.1007/s11092-015-9233-6

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford university press.

Frankel, J., Wallen, N., & Hyun, H. (2012). *How to design and evaluate research in education*. McGraw-Hill.

Gibson, W., & Brown, A. (2009). *working with qualitative data*. SAGE Publications.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Psychology Press.

Glaw, X., Inder, K., Kable, A., & Hazelton, M. (2017). Visual methodologies in qualitative research: Autophotography and photo elicitation applied to mental health research. *International Journal of Qualitative Methods,* 16(1). https://doi.org/10.1177/1609406917748215

Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation*. Sage Publications.

Hatipoğlu, Ç. (2010). Summative evaluation of an English language testing and evaluation course for future English language teachers in Turkey. *ELTED, 13,* 40-51. http://www.elted.net/uploads/7/3/1/6/7316005/v13_5hatipoglu.pdf.

Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: expectations and needs of pre-service English language teachers. *ELT Research Journal, 4*(2), 111-128. http://dergipark.org.tr/tr/pub/eltrj/issue/28780/308006

Haznedar, B. (2012). Perspectives on preservice education of English Language Teachers in Turkish primary schools. In Y. Bayyurt & Y. Bektaş-Çetinkaya (Eds.), Research Perspectives on teaching and learning: Policies and practices (pp. 39-58). Peter Lang.

 Heaton, J. B. (1990). *Writing English language tests*. Longman

Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402. https://doi.org/10.1177/0265532208090158

Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing, 30*(3), 345-362. https://doi.org/10.1177/0265532213480334

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing, 27*(4), 555-584. https://doi.org/10.1177/0265532209351431

Kaplan, R. M., & Saccuzzo, D. P. (2004). *Physchological testing: Principles, applications, and issues*. Wadsworth.

Kirkgöz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations. *RELC journal, 38*(2), 216-228. https://doi.org/10.1177/0033688207079696

Kirkgöz, Y., Babanoglu, M. P., & Agçam, R. (2017). Turkish EFL teachers' perceptions and practices of foreign language assessment in primary education. *Journal of Education and e-Learning Research, 4*(4), 163-170.

http://files.eric.ed.gov/fulltext/EJ1167153.pdf

Köksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education-TOJDE, 5*(1). Retrieved from https://dergipark.org.tr/en/pub/tojde/issue/16931/176755

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197. https://doi.org/10.1177/0265532214554321

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing, 30*(3), 329-344. https://doi.org/10.1177/0265532213480129

McKay, P. (2006). *Assessing young language learners.* Cambridge University Press

McNamara, T. (2000). *Language testing*. Oxford University Press.

Mede, E., & Atay, D. (2017). English Language Teachers' assessment literacy: The Turkish context. *Dil Dergisi, 168*(1), 1-5. Retrieved from https://www.ceeol.com/search/article-detail?id=548906

Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. Jossey-Bass

Messick, S. (1996). Validity and washback in language testing. *Language testing, 13*(3), 241-256. https://doi.org/10.1177/026553229601300302

MNE (2018). English language education curriculum. Retrieved from https://mufredat.meb.gov.tr/Dosyalar/201812411191321-.pdf

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing, 30*(3), 363-380. https://doi.org/10.1177/0265532213480336

Ölmezer Öztürk, E., & Aydin, B. (2018). Investigating language assessment knowledge of EFL teachers. *Hacettepe University Journal of Education*, 1-19. https://doi.org/10.16986/huje.2018043465

Ölmezer Öztürk, E., & Aydın, B. (2019). Voices of EFL teachers as assessors: Their opinions and needs regarding language assessment. *Journal of Qualitative Research in Education, 7*(1), 1-18. https://doi.org/10.14689/issn.2148-2624.1.7c1s.17m

Öz, H. (2014). Turkish teachers' practices on assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research, 5*(4), 775-785. https://doi.org/10.4304/jltr.5.4.775-785

Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: perceptions and practices. *ELT Research Journal, 6*, 25-44. https://dergipark.org.tr/en/pub/eltrj/issue/28779/307996.

Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*(2), 115-131. https://doi.org/10.1080/08878730009555257

Ricci, M., St-Onge, C., Xiao, J., & Young, M. (2018). Students as stakeholders in assessment: how students perceive the value of an assessment. *Perspectives on medical education, 7*(6), 352-361. https://doi.org/ 10.1007/s40037-018-0480-3

Sarıçoban, A. (2011). A study on the English language teachers' preparation of tests *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education), 41*, 398-410.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327. https://doi.org/10.1177/0265532213480128

Schoonen, R. (2011). How language ability is assessed. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 701-716). Routledge.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72*(7), 534-539. https://eric.ed.gov/?id=EJ422814

Sultana, N. (2019). Language assessment literacy: an uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia, 9*(1). https://doi.org/10.1186/s40468-019-0077-8

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Weir, C. J. (2005). Language testing and validation. *Palgrave McMillan*.

Zhang, J. (2016). Same text different processing? Exploring how raters' cognitive and meta-cognitive strategies influence rating accuracy in essay scoring. *Assessing Writing, 27*, 37-53. https://doi.org/10.1016/j.asw.2015.11.001