



Understanding language assessment literacy: Developing language assessments^{*}

Ahmet Erdost Yastıbaş^a † , Mehmet Takkaç^b 

^a Ankara, Turkey

^b Kazım Karabekir Education Faculty, Department of English Language Teaching, Erzurum, Turkey

APA Citation:

Yastıbaş, A. E., & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessment. *Journal of Language and Linguistic Studies*, 14(1), 178-193.

Submission Date:26/02/2018

Acceptance Date:08/03/2018

Abstract

Language assessment literacy has become a critical competence for a language teacher to have. Accordingly, there are many studies in the literature which have researched different aspects of language assessment literacy (i.e., language assessment training, professional development and language assessment literacy levels of language teachers). However, they have not investigated how language teachers develop appropriate language assessment according to instructional purposes. Therefore, this study has aimed to reveal the development of language assessment by language teachers. The study was designed as a qualitative study and was carried out with eight participants working in a Turkish foundation university as English language instructors. Think-aloud protocols were used to collect data and the collected data were content-analyzed. The results of the study have indicated that developing language assessments has a critical, student- and course book-centered structure which helps to make exams valid in terms of content and construct validity. This structure also helps to have positive washback effects on students. Its limitations were explained and suggestions for further studies were made.

© 2018 JLLS and the Authors - Published by JLLS.

Keywords: Language assessment literacy; developing language assessment

1. Introduction

Considering the amount of time language instructors spend on assessment and evaluation and the considerable effect of language assessment on language teaching, language instructors are given a central role in language assessment. They become the agent of language assessment who assess students and evaluates assessment data to improve students' learning and their instruction (Malone, 2013; Rea-Dickins, 2004). That is, they are in charge of every assessment-related issue. In addition, the change in educational theories creates assessment culture through which language instructors need to acknowledge education and political ideologies and social values, expectations, and attitudes by being the assessors and facilitators of language teaching (Inbar-Lourie, 2008a). Being the agent of language assessment and adapting themselves to assessment culture have gained more importance with educational and political reforms which aim to apply changes in language teaching and to prove

^{*} This article was developed based on the findings of the dissertation titled "Turkish EFL Instructors' Language Assessment Literacy: A Multiple-Case Study."

[†] Corresponding author. E-mail address: ahmeterdost@gmail.com

the changes work well by using the data language instructors provide (Brindley, 2008; Broadfoot, 2005; Inbar-Lourie, 2013; Malone, 2008; Rea-Dickins, 2008; Walters, 2010). All of these explanations make language instructors be language-assessment-literate to achieve what is expected from them. Therefore, it is essential to know language assessment literacy.

Language assessment literacy is considered as the ability that language instructors should have in order to understand, analyze and use assessment data to improve students' learning and their instruction (Inbar-Lourie, 2008b). According to Lam (2015) and O'Loughlin, (2013), language assessment literacy ability requires having the knowledge, skills and principles of assessment-related issues (i.e., test construction, use, evaluation and impact). Shortly, Malone (2008) thought that the ability is what the language instructor needs to know about language assessment and evaluation.

As the definitions of language assessment literacy ability indicate, the language instructor is supposed to have the knowledge, skills and principles of language assessment and evaluation. These knowledge, skills and principles are based on the seven standards of educational assessment proposed by the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME) and the National Education Association (NEA) in 1990. These standards involve (a) choosing assessment methods relevant to instructional purposes, (b) developing assessment tools depending on the instructional decisions, (c) administering exams, scoring them and evaluating assessment data, (d) using assessment results in decision-making related to instruction, students, school and curriculum, (e) developing valid grading procedures, (f) communicating assessment results with students and administrators and (g) recognizing illegal and unethical assessment practices (AFT et al., 1990).

1.1. Literature review

The literature review has indicated that several international and national language assessment literacy studies have been made on different aspects of language assessment literacy including the AFT and its partner organizations' standards of educational assessment. Language assessment courses were searched in some of the international studies which indicated that language assessment course instructors balanced theory and practice between 1996 and 2008 (Bailey & Brown, 2008; Brown & Bailey, 1996). However, the balance between theory and practice continued to be a problem in some language assessment courses (Jeong, 2013; Jin, 2010). The second topic studied in the international literature is professional development. Different studies showed that professional training in language assessment improved in-service language instructors' levels of language assessment literacy (Mahapatra, 2016; Montee, Bach, Donovan, & Thompson, 2013; Nier, Donovan, & Malone, 2013; Riestenberg, Di Silvio, Donovan, & Malone, 2010; Walters, 2010). Some international studies have investigated language instructors' need for training in assessment and evaluation and have revealed that language instructors needed language assessment training more (Hasselgreen, Carlsen, & Helness, 2004; Taylor, 2009; Vogt, Guerin, Sahinkarakas, Pavlou, Tsagari, & Afiri, 2008). Language instructors' assessment beliefs and practices have been investigated by some researchers including Rogers, Cheng and Hu (2007) and Munoz, Palacio and Escobar (2012) who have found that there was a disjuncture between language instructors' assessment beliefs and practices. The last topic studied in the international literature is the level of language instructors' language assessment literacy. Several studies (e.g., Kiomrs, Abdolmehdi, & Naser, 2011; Leaph, Channy, & Chan, 2015; Talib, Kamsah, Ghafar, Zakaria, & Naim, 2013; Xu & Brown, 2017) have revealed that language instructors had low or moderate levels of language assessment literacy.

In addition to the international studies, national studies also focus on pre-service assessment training. It has been found in some research that pre-service language assessment training in Turkey

was not very effective in pre-service language instructors' assessment practices (Hatipoğlu, 2010; Hatipoğlu & Erçetin, 2016). Like the results of several international studies, the language assessment literacy levels of Turkish instructors of English were low and moderate (Büyükkarcı, 2016; Hatipoğlu, 2015; Mede & Atay, 2017; Öz & Atay, 2017; Şahin, 2015). In addition, Köksal (2004) and Sarıçoban (2011) assessed and evaluated the exams prepared by Turkish teachers of English and found that the exams were improved in terms of different aspects of language assessment including validity and reliability from 2004 to 2011. Some other studies have searched language instructors' assessment beliefs, attitudes and practices and found out that though Turkish instructors of English had positive beliefs and attitudes toward different types of assessment, they could not use these types in assessing their students because of several factors including the number of the students (Büyükkarcı, 2014; Han & Kaya, 2014; Öz, 2014).

To sum up, the literature review has indicated that none of these studies has investigated the development of language assessments depending on instructional purposes. Therefore, the aim of the present study is to search this uninvestigated area by studying with Turkish instructors of English.

1.2. Research questions

The study has tried to answer the following research question.

1. How do Turkish instructors of English develop appropriate language assessments according to their instructional purposes?
 - a. What strategies do Turkish instructors of English use in developing their assessments?

2. Method

2.1. Research design

Qualitative inquiry provides insights about the phenomenon under investigation by helping to understand the underlying opinions, results and motivations related to the phenomenon (Creswell, 2007; Dörnyei, 2011). Consequently, the present study adopted qualitative research design.

2.2. Participants

Eight Turkish instructors of English working at a Turkish foundation university were chosen through purposeful sampling because purposeful sampling enabled the researchers to make their sampling line up with the purposes of the study (Creswell, 2007; Dörnyei, 2011). To choose the participants, the researchers developed seven criteria based on the standards of educational assessment proposed by AFT and its partner organizations in 1990.

Table 1. Demographic information about the participants

Participant	Gender	Age	Experience	BA	Weekly Teaching Hours	Number of Students
Instructor 1	Female	28	5	EL*	21	More than 200 students
Instructor 2	Female	35	10	ACL*	21	More than 300 students

Instructor 3	Male	30	7	ELL*	27	More than 200 students
Instructor 4	Male	28	7	ELT*	21	More than 200
Instructor 5	Male	28	5	ELT	24	More than 200 students
Instructor 6	Male	35	7	ELT	15	More than 200 students
Instructor 7	Male	30	9	ELT	21	More than 250
Instructor 8	Male	30	9	ELL	24	More than 300 students

Note: *EL: English linguistics, ACL: American culture and literature, ELL: English language and linguistics and ELT: English language teaching.

As Table 1 shows, six male and two female participants took part in the research. The participants graduated from different departments (e.g., ACL, EL, ELT and ELL). They were between 28 and 35 years old and had between five- and ten-year teaching experience. They taught English to more than 200 students between 15 and 27 hours every week.

2.3. Instrument

Think-aloud protocol helps a researcher to understand and describe what his participants focus on and how they structure what they focus during a task (Fonteyn, Kuipers, & Grobe, 1993). Therefore, the researchers in the study used think-aloud protocols with each participant to describe the cognitive processes the participants used in developing their exams. The researchers also used concurrent think-aloud protocols for having the direct verbalization of the cognitive processes and retrospective think-aloud protocols as a follow-up to have a broader picture of the cognitive processes (Fonteyn et al., 1993).

2.4. Trustworthiness

Triangulation, thick description, peer scrutiny and member checks were used to make the study trustworthy (Lincoln & Guba, 1985). First, concurrent think-aloud protocols were triangulated with retrospective think-aloud protocols. Second, the findings were thickly described. The researchers prepared the transcriptions of think-aloud protocols, developed codes and content-analyzed think-aloud protocols together. Finally, each participant checked the transcriptions and analyses of their own think-aloud protocols.

2.5. Data collection procedures

The researchers took a legal permission from the foundation university before collecting the data. Then, they informed each participant about the aim of think-aloud protocols and got the consent of each participant. A sample think-aloud protocol was made with each participant to familiarize them with think-aloud protocol procedure. Then, eight think-aloud protocols were made with each participant when all participants prepared their midterm exams. Think-aloud protocols lasted between 35 and 120 minutes. After the concurrent think-aloud protocols, retrospective think-aloud protocols were made with the participants. Think-aloud protocols were audio-recorded.

2.6. Data analysis

A recursive framework was used to content-analyze think-aloud protocols: coding, theming, organizing and interpreting (Yıldırım & Şimşek, 2013). The data were categorized into meaningful units and conceptualized through the codes given to each meaningful unit to explain the relationships. The data were read many times to code and the codes in Table 2 were used for naming them.

Table 2. Codes to analyze the think-aloud protocols

1. Starting to prepare the exams
2. Choosing reading passages, listening audio and/or words to prepare the questions
3. Deciding what to ask in the exams
4. Preparing the exam questions
5. Self-assessing the written questions
6. Evaluating the available questions to use in the new exams
7. Finishing the preparation of the exams

After the code list, the themes which covered the codes in the list were found by sorting out the similarities and differences among the codes, so the researchers could categorize the codes by placing the similar ones into a theme and explain the relationships. Consequently, the data were organized and described with the excerpts taken from think-aloud protocols. The researchers presented the data by relating them to each other without adding their comments or interpretations to the analysis. The data were interpreted without conflicting with the description of the data in the end. Then, explanations were made to make the data meaningful, to draw logical conclusions from the findings, to reveal cause and effect relationship and to show the significance of the findings.

3. Results

To understand the results, it is important to know which questions each participant prepared as listed below.

- a. Instructors 2, 4 and 6 prepared reading, vocabulary and listening questions in their midterm. Instructor 8 prepared vocabulary, listening and open-ended questions for his midterm.
- b. While Instructor 3 prepared listening and reading questions, his partner, Instructor 7 prepared vocabulary and grammar questions for their midterm.
- c. Instructor 5 was in charge of preparing vocabulary and listening questions in his midterm.
- d. Instructor 1 was responsible for preparing listening, grammar, pronunciation, reading and vocabulary questions in her midterm.

3.1. Starting to prepare the exams

Instructors 1, 3, 5 and 7 first checked what was taught in the class. Instructors 2, 4, 6 and 8 checked what was studied in the class. In addition, Instructor 8 thought about how to start his exam, Instructor 2 brainstormed about the structure of her exam and Instructors 4 and 6 decided how to prepare the exams. All participants concentrated on how to begin writing exam questions as a first or second step. Instructors 6 and 8 began with vocabulary, and Instructor 4 started with reading, Instructors 2, 3 and 5 started with listening and Instructors 1 and 7 started with grammar.

All participants had to choose a starting point. It could be reading, grammar, listening, or vocabulary. Therefore, they needed to choose listening audio, a reading passage and/or words to

prepare questions. They followed different ways to find out exam listening audio and/or reading passages. Instructor 4 used the CD of the course book and Instructor 1 used the test book and its CD of the course book. Instructors 3 and 8 benefitted from the Internet and Youtube respectively. Instructor 6 used his previous exam's reading passage and listening audio while Instructor 2 used the Internet and one of her colleagues in finding reading passages and/or listening audio for the midterm exams.

3.2. Choosing reading passages, listening audio and/or words to prepare the questions

Instructor 4 considered the similarity of the topics between the classroom reading passages and an exam reading passage, the level of the exam reading passage, and the words used in the exam reading passage in selecting a reading passage for the midterm exam as the excerpt below shows:

Instructor 4: I found a passage called neuro-marketing. Check whether its content was related to the students' department by reading fast. These topics and words are the topics and words that we always talk about in our classes. It is related to the students' department. The words are similar to the ones that we have studied in our classes, but the language used is more difficult than the one used in the reading passages that we have studied. Maybe, I can simplify the sentences, so I can use it in the exam.

Similarly, these issues were significant for Instructors 1, 2, 3 and 6. Whether the reading passages gave a lot of opportunities to prepare questions was also important for them. Instructors 2, 3 and 6 paid attention to the length of the reading passage and the time necessary for their students to read it and answer its questions.

In selecting listening audio, Instructors 1, 2, 3, 4, 6 and 8 took into account the students' levels of English and the similarity between the chosen audio and the classroom listening audio in terms of topic. These participants also took into account that the audio was understandable by the students and that it had a clear and audible recording. Instructors 2, 3 and 6 took into consideration the length of the audio. Instructors 2, 4 and 8 controlled whether the audio included the words that they taught in the class. Besides, being able to prepare questions from the audio was significant for Instructor 4 in selecting the midterm audio. Instructor 5 wrote his own listening script, while the other participants preferred to select the exam audio from the websites or CDs. Instructor 5 selected two of the topics studied in the listening parts of his classes and believed he could integrate with each other easily to write his own script. He also followed the points which the other participants did in selecting the midterm audio from the CDs and websites. In addition, he sometimes had difficulties in preparing a question from the script, so he had to make changes in the script and enhanced his script.

Instructors 5, 7 and 8 chose the words they emphasized a lot in the class in selecting words to prepare vocabulary questions. In addition, these participants made personal judgments about the words in terms of whether the words were easy or difficult and whether they liked the words. Instructors 5 and 8 randomly chose the words from the course book exercises. Instructor 8 also selected the words which he found tricky.

3.3. Deciding what to ask in the exams

Instructors 2, 3, 4, 5 and 8 had to decide what to ask from listening audio and/or a reading passage in the midterm exam. These participants first listened and/or read. During reading and/or listening, they considered a piece of information significant, so they prepared a question for that piece of information. They sought to find a piece of information with which they could assess a certain listening and/or reading skill. They also tried to find a piece of information for which they thought they could prepare a question. The excerpt below clearly reveals this finding:

Instructor 3: I am going on writing the second question. Now, there are pieces of general information about the life of Jason Stone, the place he died and how old he was in the part that I have listened. The ones which have caught my attention most among these pieces of information in the part I have listened are the place where he died, how old he was when he died and his constant business trips. They have caught my attention a lot. I have to empathize my students when I listen to something. When I listen, I have to determine which part my students can understand better and where the speaker emphasizes a piece of information. Therefore, my second question will be about where he died. My second question is “Where did he die?”. It is in his London home. When I look at the previous answer, it is the option b, so I am thinking of writing the correct answer in the option b. When I say London home here, where could he die? Which place comes to my mind? It might be a place in the house. It might be the working room or office. I should mention them especially because I have asked my first question related to his job. I can use something related to his job as a distractor. Therefore, I will write in his office in the option a. I wrote the correct answer in the option b. I will write in his study which is completely unrelated in the option c. My aim is to check whether my students can listen for finding specific information to answer the question.

In addition, Instructor 8 determined what to ask by taking into account his students’ attitudes toward the parts of his lessons. His students did not pay enough attention to the questions in the get ready parts of his course book, so he decided to ask such questions in his midterm exam.

3.4. Preparing the questions

The participants preferred to use one or two of the following ways: writing original questions and using the available ones. The participants’ choices were explained below.

- a. Instructor 1: She selected midterm exam questions from the test book of the course book by only adding the fourth options to the questions.
- b. Instructor 2: She used the available listening questions on the website and also wrote her own listening questions. She evaluated the reading questions prepared by her colleague and used them with some changes. She used the vocabulary questions of her previous midterm exam.
- c. Instructor 3: He wrote his own listening and reading questions.
- d. Instructor 4: He prepared his own reading and listening questions. He used the vocabulary part of his previous midterm exam.
- e. Instructor 5: He wrote his own listening questions. He wrote some of his vocabulary on his own, developed some of them based on the course book exercises and used some of the course book exercises as the exam questions without changing.
- f. Instructor 6: He used his previous midterm exam without making any change.
- g. Instructor 7: He wrote most grammar and vocabulary questions on his own and used a dictionary and grammar book to write the other questions.
- h. Instructor 8: He prepared vocabulary and open-ended questions by using the exercises in his course book without changing or with some changes. He also wrote some of his vocabulary questions and listening questions by himself.

All participants continuously brainstormed and outlined during writing original questions. They determined the number, types, content and timing of the questions as well as the weights of different

skills or sections in the exams in their brainstorming and outlining. The quotation below clearly exemplifies these procedures:

Instructor 5: How many questions can I ask from this dialogue? I will check how many questions I can ask from this. Actually, there is not a limitation on the number of the questions. There may be three questions or five questions. However, I will try to ask as many questions as possible from the dialogue. OK! This is the topic. There is a product. It is a defect one and causes a problem. What type of questions can I use here? I may use multiple-choice and true-false questions. Let's start with two true-false questions.

Instructors 3 and 4 also brainstormed and outlined the number of the options for listening and reading questions. Both participants preferred using three options with listening questions and four options with reading questions because they knew that their students were not good at listening and did not have good scores, so they wanted to make the listening questions easier than the reading questions.

Instructors 2, 3, 4, 5, 7 and 8 wrote all, most, or some of their exam questions by themselves. First, they talked to themselves about what to ask. Second, they brainstormed about the content of the exams in their private speech. Then they code-switched in writing the stem, the options, or both. The following excerpts clearly indicate these procedures:

Instructor 4: Our first question is generally what the passage is about. What is the passage mostly about? The passage generally mentions neuro-marketing. It is finding out the clients' brand choices by obtaining their reactions in their brain when they see brands related to a type of product through placing electrodes on their heads. I am writing about this. Its correct answer is, a, a new method to learn consumer choices. Generally, a new way or method of learning consumers' choices.

Instructor 7: The second one is to encourage. What can I write for it? I mentioned the structure 'encourage someone to do something' and made them [his students] write their own sentences. Therefore, I should definitely ask it. Let's do it like this. Their departments are related to teaching. Therefore, I should write a sentence related to being an effective Instructor. An effective Instructor should --- their students to participate to... Is participate used with to or in? Participate to or participate in? Yes, an effective Instructor... They can learn a feature of an effective Instructor. An effective Instructor encourages his students to participate in classroom activities actively. This is a good one.

Instructors 2, 3, 4, 7 and 8 often referred back to what they had taught in the class in writing their own questions. These participants including Instructor 5 used the types of the questions (like matching, fill-in-the-blank, true-false and multiple-choice) which were similar to their classroom activities and which their students were familiar with. They paid attention to whether the students could understand and answer the questions in writing their own questions. Similarly, Instructor 2 considered her students' comments about her previous exams, so she sought not to prepare her questions in the way that her students had complained about. Instructor 8 took into account his students' motivation in developing his questions. Like Instructor 2, Instructor 5 tried not to ask any question about which his students might complain after the exam. In spite of this, these participants also rendered their exams challenging enough to check who studied and did not study. Instructors 2, 3, 4, 5, 7 and 8 related the content of the questions to the topics taught in the class. Instructor 7 also related his questions to the students' future professions and daily lives.

Instructor 5 concentrated on the lengths of the options when he prepared them. Instructors 3 and 5 believed that there would be two options very close to each other in terms of the correct answer, one

option not related to the correct answer and the last option that was not related to the correct answer, but close to it.

Instructors 2, 3, 4, 5 and 8 ordered listening and reading questions according to the order of the events in the audio and/or reading passage. Instructors 3 and 4 preferred to paraphrase the reading questions' options. Instructors 5, 7 and 8 randomly ordered vocabulary questions. In writing vocabulary questions, Instructors 5 and 7 used the same parts of speech in the vocabulary questions' options.

Instructors 5, 7 and 8 wrote their own vocabulary questions. They wrote their own sentences in some vocabulary questions and used other ways in other vocabulary questions. Instructor 7 used a dictionary to write the definitions of the matching questions and multiple-choice fill-in-the-blanks questions. Instructor 5 benefitted from his classroom examples and the examples in the course book in writing the stems of the multiple-choice vocabulary questions. Instructors 5 and 8 used the words and their definitions from the matching exercises in the course books as their midterm matching questions without changing. In addition, Instructor 8 also chose and used some other vocabulary exercises in the course book as his midterm vocabulary questions either without or with changing. He also converted the definition of a word in a matching exercise of the course book into a fill-in-the-blank question in the midterm exam.

3.5. Self-assessing the written questions

Instructors 2, 3, 4, 5, 7 and 8 self-assessed the questions after preparing one of or all of the questions. In self-assessing the questions, these participants checked whether the stems could be understood by the students, whether the wording and the use of grammar were correct in the stems and the options, whether the questions could assess what they wanted them to assess and whether the answers of the questions were prepared correctly. These procedures are clearly understood from the excerpt below:

Instructor 5: For example, I can ask it. It can be similar to the example that I gave in the class. For instance, a famous singer may sue against a newspaper. Why? Because of law... For example, I can say "Every day we read in the newspapers that one of the celebrities, celebrities, every day we read in the newspapers that one of the celebrities sue against, one of the celebrities sues against a." We read in the newspaper that he/she sued against a newspaper. This sounds a little weird. Or we can say we hear. Every day we hear that. We hear that one of the celebrities sues against a newspaper or magazine, a magazine or newspaper, because of ... What is the correct answer? What should I say in the option a? Or where should I write the correct answer? We have four options: a, b, c and d. For example, I should write the correct answer in the option b. Deformation, blackening someone. He/she sued because of deformation. The answer is deformation. We chose and asked a noun. We should use nouns in the other options. For example, we use intent meaning willingness. What else can I use? What else can I use? Let's look at the other units. We can use notion. I used notion because it is a noun. Another noun? I can use movement. Let's check other options whether they can also be answers. Movement... Because of the person who wanted, he/she sued. He/she sued because of deformation. The right answer is b. It cannot be the answer that he/she sued because of intention and notion.

3.6. Evaluating the available questions to use in the new exams

Instructors 2, 4 and 6 decided to use the vocabulary questions of the previous midterm exams in the new midterm exams. First, they checked whether they used the same syllabi and course books in the

previous term. Second, they matched the units for which the old vocabulary questions were prepared with the units for which they would prepare new vocabulary questions. They considered whether they had experienced any problem with the questions in the previous midterm exam, which was also effective in their decisions. As the vocabulary questions in the previous midterm exams met these criteria, these participants decided to use the previous questions.

Like Instructors 2, 4 and 6, Instructor 2 also decided to use the reading questions which one of her colleagues had given to her and the available listening questions on the Internet. She checked whether her students could understand and answer those questions. Once she understood that the questions were understandable and answerable for the students, she decided to use them in the midterm exam.

Similarly, Instructor 6 decided to use the listening and reading questions of his previous midterm exam. He self-assessed the listening audio and reading passage in terms of the criteria mentioned in the paragraph where how exam listening audio and reading passages were chosen to prepare listening and reading questions was explained. Then, he checked the similarity between the listening and reading questions and classroom activities as well as the understandability and answerability of the previous exam questions for his students. Once he decided that the questions were relevant to use in the new midterm exam, he used the questions without changing. He did so because of the number of the students and lack of time.

In addition to these participants, Instructor 1 selected the midterm questions from the test book of her course book because she did not believe that she was an expert on preparing questions. It was also because selecting questions among the available one was time-saving, using the test book of the course book provided content-validity and she did not experience any problem with the course book, its exercises and its answer keys before. Besides, she considered testing environment, testing program (Blackboard) and the students' levels of English. She selected the questions related to what she had taught in the class. She chose the exam listening audio and reading passage whose topics were similar to the ones used in the course book. She also added one more option to the questions to make them more challenging.

3.7. Finishing the preparation of the exams

All participants self-assessed the questions again when they finished preparing the exams. The participants who prepared reading and/or listening questions also assessed the chosen exam reading passages and/or listening audio. These participants considered the duration, recording, topic and understandability of the listening audio when self-assessing the audio. The quotation below exemplifies how the chosen exam listening audio and the questions prepared in the midterm exams were self-assessed:

Instructor 5: The things that we pay attention to in preparing listening questions is finding a related listening audio. It is suitable for their [his students] levels, easy for them to understand, related to classroom topics and includes a lot of words related to the classroom topics. Besides, the questions I prepared from the listening audio are the question types that my students have practiced in my classes and in the course book and that they are familiar with.

The participants considered whether the chosen exam reading passages included the words taught in the class, whether the passages were understandable for the students and whether the passages were similar to the classroom reading passages in terms of length and topic. Besides, Instructor 3 assessed his instructions and the question words used in the questions to be sure that the students could understand the questions. Instructor 2 and 8 also self-assessed the variety of the questions in the exams.

In addition, some participants prepared the answer keys while writing the questions, whereas the others prepared the answer keys after preparing the questions. All participants checked the number of the questions prepared for each skill and wanted to ensure that the numbers reflected the weights given to the different skills in the class. They also scored the questions in the end. Instructors 1, 3, 4, 5, 6 and 7 scored the exams depending on the number of the questions in the exams. Instructor 2 gave scores to the questions according to the weights of English for academic purposes and English for specific purposes in her three-hour academic English course. Instructor 8 scored his questions based on their difficulty levels.

4. Discussion

The results of the study have first indicated that the participants are very dependent on the course books when they prepare their exams. That is, the course books are very influential in deciding how to choose reading passage and listening audio for the midterm exam, what types of questions should be used and how to develop questions and their content and how to evaluate and choose questions among the available ones. This result has revealed that the Turkish instructors of English in the study considered content validity in preparing exam questions. The finding conflicts with the findings of Köksal (2004) and Sarıçoban (2011) which revealed that the exams prepared by the Turkish teachers of English had content validity problem. In addition to this problem, these researchers also found that those exams were problematic in terms of construct validity. However, the Turkish instructors of English in this study built the construct of their midterm exams based on their course books. The questions the instructors prepared really assessed the determined construct, which means the Turkish instructors of English in the study improved their exams in terms of construct validity. This finding may result from the fact that the instructors in this study were course-book centered.

The results of the study have secondly demonstrated that the Turkish instructors of English considered their students most in developing their exams. For instance, the instructors referred back to what they had taught in the class in writing and choosing exam questions and paid attention to the students' comments and levels of English in choosing exam reading passage and listening audio and writing options. This finding must be closely related to the instructors' beliefs about language teaching because it is known that student-centered teachers consider students most in teaching and assessment. As Davison (2004) and Scarino (2013) stated, teachers' beliefs influence their assessment practices. Similarly, the result of the study has pointed out that some Turkish instructors of English decided the number of the options in listening and reading questions depending on their beliefs.

The third important result is that the Turkish instructors of English in this study paid attention to the washback effects of their midterm exams. As aforementioned, the instructors were course book-centered, so they made the midterm exams content valid, aligned the midterm exams with their goals and objectives and reflected the classroom activities with the exams to create positive washback effects on the students. These ways are also suggested in the literature by Brown (2004) and Rogier (2014). It also seems that the instructors' being student-centered in language teaching contributed to creating positive washback effects because being student-centered requires doing every instructional and assessable activities for the sake of students.

The fourth and most important result of this study is that the Turkish instructors of English in this study adopted a critical attitude toward each phase of developing their midterm exams. To exemplify, the instructors developed several certain criteria to choose reading passage and listening audio for the midterm exams, to evaluate and select questions among the available ones and to self-assess their questions. The instructors also used a few critical thinking strategies (e.g., brainstorming, outlining

and private speech) in writing their questions. This finding corroborates several researchers (e.g., Inbar-Lourie, 2008b; La, 2015; Malone; 2008; O’Loughlin, 2013) who mentioned that language assessment literacy requires being critical in every assessment-related activity. According to the results of the study, self-assessment was the most frequently used strategy in developing exams, which most probably contributed to the instructors’ being critical because self-assessment helps to identify strengths and weaknesses and overcome weaknesses through finding solutions on one’s own (Takkaç Tulgar, 2017).

5. Conclusions

The literature review has indicated that there is not any study focusing on how language teachers develop appropriate language assessments according to their instructional purposes. Understanding this procedure is necessary for understanding the implementation of language assessment literacy. In accordance with this finding, the present study indicates that developing language assessments has a critical, student- and course book-centered structure. Development language assessments is critical because it requires using brainstorming, outlining, private speech and self-assessment in the pre-, while and post-exam preparation processes. It is student-centered since students are considered most in every assessment-related activity. It is course book-centered because course books are the key determiner of selecting reading passage and listening audio for the exams, the types of the questions to prepare, self-assessing questions and evaluating available questions to use in the exams. This critical, student- and course book-centered structure can enable teachers of English to make their exams valid and have positive washback effects on the students.

The study is first limited because it has a small sampling size as a result of its qualitative nature and has been conducted in a foundation university setting. The second limitation is that the study has investigated how formal, summative and selected-response exams have been prepared. Consequently, it is recommended for further studies that similar studies should be made in other settings (e.g., private and public primary, middle and high school as well as public university) to understand how Turkish teachers/instructors of English working in these settings develop appropriate language assessments depending on their instructional goals. Secondly, the study also suggests further studies research how other assessment methods (e.g., constructed response, performance assessment and personal communication) are developed by Turkish teachers/instructors of English for formal, informal, summative and/or formative assessment purposes.

Acknowledgements

This article is developed depending on the data of the dissertation titled “Turkish EFL Instructors’ Foreign Language Assessment Literacy: A Multiple-Case Study” which is supported by 2211-A Scholarship Program of TÜBİTAK.

References

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Retrieved from <http://buros.org/standards-teacher-competence-educational-assessment-students>

- Bailey, K. M. & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236-256). Clevedon, UK: Multilingual Matters.
- Brindley, G. (2008). Educational reform and language testing. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.) (pp. 365-378). Springer Science+Business Media LLC.
- Broadfoot, P. M. (2005). Dark alleys and blind bends: Testing the language of learning. *Language Testing*, 22(2), 123-141. doi: 10.1191/0265532205lt302oa
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Turkey: Pearson Education Inc.
- Brown, J. D. & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349-383. doi: 10.1177/0265532208090157
- Büyükkarcı, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction*, 7(1), 107-120.
- Büyükkarcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Eğitim ve Öğretim Dergisi*, 6(3), 333-346.
- Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, California, the United States of America: Sage Publications.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334. doi: 10.1191/0265532204lt286oa
- Dörnyei, Z. (2011). *Research methods in applied linguistics*. Oxford, UK: Oxford University Press.
- Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research*, 3(4), 430-441.
- Han, T. & Kaya, H. İ. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education*, 4(1), 77-93.
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs. Report: Part one – general findings*. Retrieved from <http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>
- Hatipoğlu, Ç. (2010, Winter). Summative evaluation of an English language testing and evaluation course for future English language teachers in Turkey. *ELTED*, 13, 40-51.
- Hatipoğlu, Ç. (2015, May). *Diversity in language testing and assessment literacy of language teachers in Turkey*. Paper presented at the 3rd ULEAD Congress, International Congress on Applied Linguistics: Current Issues in Applied Linguistics, Çanakkale, Turkey.
- Hatipoğlu, Ç. & Erçetin, G. (2016). Türkiye'de yabancı dilde ölçme ve değerlendirme eğitiminin geçmişi ve bugünü. In S. Akcan & Y. Bayyurt (Eds), 3. *ulusal yabancı dil eğitimi kurultayı: Türkiye'deki yabancı dil eğitimi üzerine görüş ve düşünceler 23-24 Ekim 2014, konferanstan seçkiler* (pp. 72-89). İstanbul, Turkey: Boğaziçi University.
- Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. doi: 10.1177/0265532208090158

- Inbar-Lourie, O. (2008b). Language assessment culture. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.) (pp. 285-299). Springer Science+Business Media LLC.
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-8). Blackwell Publishing Ltd.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345-362.
- Jin, Y. (2010). The place of language testing and assessment in the preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584. doi: 10.1177/0265532209351431
- Kiomrs, R., Abdolmehdi, R., & Naser, R. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Testing*, 4(1), 156-161.
- 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).
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197. doi: 10.1177/0265532214554321
- Leaph, K., Channy, M., & Chan, N. K. (2015). Cambodian ELT university practitioners' use of standardized tests for practice and assessment. *Language Education in Asia*, 6(1), 4-16.
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.
- Mahapatra, S. K. (2016). Tracking the development of teachers' language assessment literacy: A case study. In G. Pickering & P. Gunashekar (Eds.), *Ensuring quality in English language teacher education* (pp. 106-114). New Delhi, India: British Council.
- Malone, M. E. (2008). Training in language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.) (pp. 225-239). Springer Science+Business Media LLC.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344. doi: 10.1177/0265532213480129
- Mede, E. & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Ankara Üniversitesi TÖMER Dil Dergisi*, 168(1), 43-60.
- Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013, Spring). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 14, 1-32.
- Munoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *PROFILE*, 14(1), 143-158.
- Nier, V. C., Donovan, A. E., & Malone, M. E. (2013). *Promoting assessment literacy for language instructors through an online course*. Retrieved from http://www.cal.org/ecolt/pdfs/AB_Poster_10-21-2013_FINAL.pdf
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380. doi: 10.1177/0265532213480336

- Öz, H. (2014). Turkish teachers' practices for *assessment for learning* in the English as a foreign language classroom. *Journal of Language Teaching and Research*, 5(4), 775-785.
- Öz, S. & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249-258.
- Rea-Dickins, P. (2008). Classroom-based language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.) (pp. 257-271). Springer Science+Business Media LLC.
- Riestedberg, K., Silvio, F. D., Donovan, A., & Malone, M. E. (2010, Fall). Development of a computer-based workshop to foster language assessment literacy. *Journal of the National Council of Less Commonly Taught Languages*, 9, 21-42.
- Rogers, W. T., Cheng, L., & Hu, H. (2007). ESL/EFL instructors' beliefs about assessment and evaluation. *Canadian and International Education/Education canadienne at internationale*, 36(1), 39-61.
- Rogier, D. (2014). Assessment literacy: Building a base for better teaching and learning. *English Teaching Forum*, 3, 2-13.
- Şahin, S. (2015, May). *Language testing and assessment (LTA) literacy of high school English language teachers in Turkey*. Paper presented at the 3rd ULEAD Congress, International Congress on Applied Linguistics: Current Issues in Applied Linguistics, Çanakkale, Turkey.
- Sarıçoban, A. (2011). A study on the English language teachers' preparation of tests. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 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. doi: 10.1177/0265532213480128
- Talib, R., Kamsah, M. Z., Ghafar, M. N. A., Zakaria, M. A. Z. M., & Naim, H. A. (2013). *T-assess: Assessment literacy for Malaysian teachers*. Paper presented at the International Conference on Assessment for Higher Education Across Domains and Skills, Kuala Lumpur.
- Tulgar, A. T. (2017). Selfie@ssessment as an alternative form of self-assessment at undergraduate level in higher education. *Journal of Language and Linguistics Studies*, 13(1), 321-335.
- Vogt, K., Guerin, E., Sahinkarakas, S., Pavlou, P., Tsagari, D., & Afiri, Q. (2008). *Assessment literacy of foreign language teachers in Europe – current trends and future perspectives*. Paper presented at the 5th EALTA Conference, Athens, Greece.
- Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, 7(4), 317-342. doi: 10.1080/15434303.2010.516042
- Xu, Y. & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.
- Yıldırım, A. & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri* (9th ed.). Ankara: Seçkin Yayıncılık.

Yabancı dilde ölçme değerlendirme okuryazarlığını anlamak: Dil değerlendirilmeleri geliştirmek

Öz

Yabancı dilde ölçme değerlendirme okuryazarlığı bir dil öğretmenin sahip olması gereken önemli bir yeterlik olmuştur. Bunun sonucunda yabancı dilde ölçme değerlendirme okuryazarlığının farklı boyutlarını (yabancı dilde ölçme değerlendirme eğitimi, profesyonel gelişim ve dil öğretmenlerinin ölçme değerlendirme okuryazarlık seviyesi gibi) araştıran çok fazla çalışma alinyazında bulunmaktadır. Bununla birlikte, bu çalışmalar dil öğretmenlerinin eğitimsel amaçlara uygun dil değerlendirmesini nasıl geliştirdiklerini araştırmamıştır. Bu yüzden bu çalışma, dil değerlendirmelerinin dil öğretmenleri tarafından nasıl geliştirildiklerini göstermeyi amaçlamıştır. Çalışma, nitel bir araştırma olarak tasarlanmış ve bir vakıf üniversitesinde çalışan sekiz Türk İngilizce okutmanı ile yürütülmüştür. Sesli düşünme protokolleri very toplamak için kullanılmış ve kullanılar veriler içerik analizi kullanarak analiz edilmiştir. Çalışmanın sonuçları; dil değerlendirmeleri geliştirmenin, eleştirel, öğrenci ve ders kitabı merkezli bir yapıya sahip olduğunu göstermiştir ki bu yapı, sınavları içerik ve yapı açısından geçerli kılmaya ve sınavların öğrenciler üzerinde olumlu etkilere sahip olmasına yardımcı olmaktadır. Çalışmanın sınırlılıkları açıklanmış ve sonraki çalışmalar için çeşitli önerilerde bulunulmuştur.

Anahtar sözcükler: Yabancı dilde ölçme değerlendirme okuryazarlığı; dil değerlendirmesi geliştirmek

AUTHOR BIODATA

Ahmet Erdost YASTIBAŞ: He has a Ph.D. degree in English language teaching and his research interests include language assessment and evaluation.

Mehmet TAKKAÇ: He has a Ph.D. degree in English language teaching and his research interests include teacher training and professional development.