



TITLE: TRANSFORMING LANGUAGE ASSESSMENT: A CASE STUDY ON ENHANCING LISTENING SKILLS EVALUATION FOR LANGUAGE LEARNERS

Author: Dilrabo Babakulova

Workplace: Westminster International University in Tashkent

Position: Associate Lecturer

Abstract

This study examines the transformation of language assessment through a case study focused on enhancing listening skills evaluation for language learners. Critiquing the original assessment's effectiveness in terms of practicality, reliability, validity, authenticity, and washback, the article proposes modifications to improve its efficacy. The modified assessment includes lead-in questions, detailed instructions, and a reflective writing task, aligning with established testing principles. Findings indicate that these modifications enhance task authenticity, comprehension, and student engagement, ultimately leading to a more effective assessment process.

Key words: Language assessment, test modification, test analysis, task authenticity, test validity and reliability

Introduction

Assessment is one of the crucial features of language teaching which is vitally important to see how much knowledge learners could acquire through the course. Moreover, by assessing it is possible to identify the challenges that students are facing and compare the expected and actual results of learners. According to Brown (2010) "assessment is an ongoing process that encompasses a wide range of methodological techniques (p.3)". Assessment is very wide understanding that is why it is important to identify its types. There are 2 forms of assessment that are used to evaluate learners' understanding and performance of language skills. Formative assessment is "assessment for learning" that is ongoing process which provides feedback on learners' progress during the course. Summative assessment is used to measure learners' improvement at the end of the course. This type of assessment is "assessment of learning".

This project paper analyzes one of the summative assessments that was used for level 3 students of Westminster Academic Lyceum to evaluate their listening skills as a midterm test. Moreover, it presents modified test version also as well which can be used further with the aim of getting more reliable results.

Learner Assessment Profile

In order to compare original and modified versions of listening test according to 5 principles of testing, a volunteer learner was invited to participate. As the test was designed and was already used for students of Academic Lyceum under Westminster University before, I have chosen a second year student of the same site to analyze those tests.

Anvar (the name was changed for privacy purposes) is 17 years old student at Academic Lyceum under Westminster International University in Tashkent. He is Uzbek, knows Russian



as a native speaker; moreover, he speaks English at B2 level and now he learns Spanish as a foreign language. First, he started taking English courses at the age of 9-10 individually where he acquired basic knowledge of the language grammar and learnt some common words. After 5 years he realized that his speaking wasn't fluent enough since he had difficulties in expressing his opinion; that is why he continued his studies at the learning center by joining to the group for pre-intermediate level learners. He enjoyed studying at the language learning center where competition was involved among learners. So that he improved his speaking skills and studied till he finished upper intermediate level. Moreover, he took a 2-month IELTS course; however, he left without finishing this course as his teacher left the work.

In summer 2018 he took an exam from English and mathematics to Westminster lyceum and passed successfully by gaining 64.5 points. The English exam was based on reading, multiple-choice grammar questions and candidates were asked to produce 2 writing pieces. The examination for checking knowledge of English wasn't difficult for him and he gained 83 out of 100 points. Nevertheless, he had challenges with the math test, and could solve only 43 % of the exam. His good language skills were the main reason for becoming a student of one of the prestigious lyceums in Uzbekistan.

In 2019, March he took IELTS exam at British Council and got 7 band score overall by hitting 8 from listening, 6.5 for speaking, 7.5 from reading and 6.5 for writing skills.

Although he is a student of Westminster academic lyceum, he took a 3-month English course, "English for Success" which was offered by American Embassy during his summer vocation.

In order to study in this course, he took a test and was able to be one of the 20 students who were chosen among 4000 candidates. In this course, he was able to improve his speaking competence; in addition, his writing skills progressed a lot. Besides advancing language skills, Anvar could broaden his common knowledge about the world and made friends.

Westminster lyceum offers a great opportunity to students to practice their English at high level since it is 3 times a week and lessons are held only in a target language by 3 different instructors. Sessions are divided according to aspects and skills of language, for instance, if one lesson is given for teaching writing and listening, the second is devoted to reading and speaking, and the last session of the week focuses on teaching grammar. Learners study English 120 academic hours in a semester; furthermore, they have 3 summative assessments: 2 midterm exams and one final test.

As students enter the lyceum by taking exam and their level of English is almost the same, learners are not given a test to place them into level groups. Most of the students are ambitious, so they already get IELTS certificate till they graduate from lyceum. Moreover, they have a chance to become Westminster University student with their grades from lyceum. That's why, students study hard in order to get high results from main subjects and exams.

The ways of teaching languages, especially English and assessing knowledge of students at Westminster lyceum is different from other public institutions as it is under international university. According to the grading system of lyceum, students acquire their overall score from formative and summative assessments. Moreover, 50% of their grades come from summative assessment which is based on mid-term and final exams. In most of the cases, mid-term and final exams are paper-based test which consist of questions to check listening, reading, writing skills, and knowledge of grammar and vocabulary aspects. In

listening section of the test, students listen to the recording, sometimes take notes and answer to true/false, multiple-choice, matching, gap-filling test type questions. In reading they are required to read the text and respond to given questions. Then, in writing learners are supposed to compose essay on given topic. In grammar and vocabulary parts, students are given mostly multiple choice questions to complete the task. Although testing system of this lyceum is a model for other educational organizations, exams also contain some drawbacks as well.

Critique of existing test and assessment

The test which was designed for 3rd level students of Westminster lyceum was prepared by English language teachers; furthermore, it was approved by the head of “Foreign languages department” and deputy of director. The original test contains 3 parts including questions to check reading and listening skills, and knowledge of language vocabulary. In the reading section a text with multiple-choice and matching questions were given, in the vocabulary part learners were mostly supposed to identify the meaning of covered vocabulary in context, and in the listening part students first had to take notes while listening to the recording and then answer to questions with the help of their notes. In mid-term exam, marks for each section were divided: 30 points for reading, 30 points for listening, and 40 points for vocabulary parts. In the listening section, learners could gain 7 points for their notes, 15 points for matching people with their ideas, and 8 points for multiple-choice questions.

When the test was used, many of the participants had problems in understanding the listening extract and taking notes while listening as it was an authentic radio-talk. They just wrote some chunks from listening and randomly answered to multiple-choice and matching questions. Moreover, test-takers also complained that it was difficult to understand listening as there was distracting background noise. For these reasons,



the listening part of the exam was modified which wasn't successful when it was first administrated among lyceum students.

According to Brown (2010) an appropriate and useful test should include following 5 principles: practicality, reliability, validity, authenticity, and washback. That is why any test should be analyzed according to those principles of testing before administering it in order to get right and positive results.

Bachman (1996) states that a test is supposed to be easy in terms of designing, administering, scoring and getting the results which refers to practicality of testing. Practicality deals with management of time for developing the test and grading it, administration of budget for taking the test and retaking it if needed and organization of feedback delivery. Taking into consideration all these features of practicality it is obvious that the listening test is not very practical, since in the first part of the listening task students take notes and the teacher is supposed to read them and score. However, the second part of the test is practical as it is easy to develop; furthermore, it takes a little time to check and score since 13 items consist of only matching and multiple choice question types where students just write correct letters in their answer sheet. Moreover, it doesn't cost a lot since only one paper is needed for each student; in addition, the cost for paper and printing is covered by the lyceum.

Reliability is following major principle in contributing to a valuable language assessment. Brown (1996) claims that "In general, the test reliability is defined as the extent to which the results can be considered consistent or stable. (p. 192)". Moreover, reliable tests should have understandable explanation in evaluating and contain a rubric for scoring (Brown, 2010). Regarding to listening test, it might be considered half reliable test as most of test takers got the same score; on the other hand, instructions were not given properly for note-taking part.

Brown (1996) states that "Test validity is defined here as the degree to which a test measures what it claims, or purports, to be measuring (p. 231)". In terms of analyzing validity of listening test, we can identify that it is not appropriate and meaningful to score their ability to take notes since in the exam students listened to radio talk while during the lessons from time to time they practiced the same task by listening to university lectures. Moreover, the students who were unable to take notes could easily answer to questions randomly as the test contained multiple-choice questions.

Advance Consulting for Education (2011) defines authenticity as type of test item and material that is related to a real-world. So, we can divide this principle into 2 categories as task authenticity and material authenticity. Although listening extract is authentic, students will not use the skill of guessing correct answer from multiple choice questions in a real life.

The last significant principle of testing is washback which is defines as having positive influence both on teaching and learning by Brown (2010). Moreover, washback may have negative impact also if learners are unmotivated after accomplishing the test and getting inappropriate results. Concerning to listening task, we can observe that learners just took some notes, maybe even meaningless words and randomly answered to multiple choice questions. Most of them were not satisfied with the recording and test items during the test, and as a result the participants were unenthusiastic in learning. That is the reason why it is hard for the teacher to provide appropriate feedback on students learning progress.



Original version of the test:

Listening (Total: 30points)

Listen to National Public Radio's Talk. Margot Adler interviews people about how much news we need. Take notes, pay attention to 5 people's opinion about news and use your notes to answer questions (Task I and Task II) on the next page. (7 points for taking notes)

Task I (15 points)

1. Andrew Weil (doctor, author)	a. Novels are more interesting than news
2. Mark Harris (author)	b. It's better not to read the news every day.
3. Gabrielle Spiegel (Johns Hopkins University)	c. People should gradually reduce their news intake.
4. John Sommerville (author, history professor)	d. The news focuses too much on problems and doesn't offer solutions
5. Tupton Shudrun (Buddhist nun, teacher)	e. Fantasy and humor are more important than the news.

Task II (8 points)

6. Mark Harris wrote an essay in the <i>New York Times</i> in the early 1970s about newspapers. Since then, he has _____.
a. changed his opinion completely
b. adjusted his opinion a little
c. maintained the same opinion
7. In Harris's opinion, novels get readers to focus on interesting people, such as _____.
a. sports figures who don't win
b. teachers at small universities
c. historical figures who are little known



- | | |
|-----|---|
| 8. | The period in history that Gabrielle Spiegel studies is the ____ centuries. |
| a. | 13 th and 14 th |
| b. | 15 th and 16 th |
| c. | 19 th and 20 th |
| 9. | Spiegel doesn't read newspapers because _____. |
| a. | they are repetitive |
| b. | she doesn't have enough time |
| c. | she wants to spend time with her children |
| 10. | In Sommerville's opinion, daily news _____. |
| a. | gives too much information |
| b. | keeps people from seeing connections |
| c. | should focus more on religion |
| 11. | Sommerville reads the news _____. |
| a. | every week |
| b. | every three or four months |
| c. | almost never |
| 12. | Tuption Shudrun is critical of the media because it _____. |
| a. | doesn't explain problems well |



- b. creates a sense of despair
- c. is difficult to understand

13. Gabrielle Spiegel thinks that children _____

a. can't understand much of news they hear

b. need to spend more time alone

c. should listen to the news with their parents

Answer box:

1	2	3	4	5	6	7	8	9	10	11	12	13

Modified version of test assessment

It is common to use the terms of “assessment” and “tests” interchangeably, yet they are different. Tests are part of assessment which measures only one field of knowledge. Tests should be used as one of the tools to assess learners’ understanding. Brown (2010) states that “tests are a subset of assessment, but they are certainly not the only form of assessment that a teacher can make”. Although tests can be useful devices, they are only one among many procedures and tasks that teachers can ultimately use to assess (and measure) students” (p. 5). However, in most of the cases, tests have some negative washback on students’ learning that constantly scare them. Therefore, it is vital to provide a test which follows 5 testing principles in order to get successful results.

By modifying the test, I tried to make it more task-authentic, reliable and valid test which offers a positive washback by assessing students work according to the criteria that will be provided beforehand. The first step in modification was including lead-in questions which will be helpful to learners to get ready to listen to the radio talk by getting general knowledge of the topic. Note taking task remained as Hughes (2003) states that “For extended listening, such as a lecture, a useful first step is to listen to the passage and note down what it is that candidates should be able to get from the passage. (p. 164)”. However, this part was a bit changed by providing the names of speakers and one of their ideas about news since



simultaneously listening and note taking was challenging for test takers when it was first used. In order to skip subjectivity in checking notes for listening task, a rubric (Appendix 1) was introduced to students to make them aware of scoring system. Furthermore, the points given for note taking were increased as it is also graded according to certain criteria. Multiple choice questions stayed the same; on the other hand, total score for this task was decreased as it doesn't check comprehension directly. As mentioned above, the lyceum English language mid-term test was designed to check only 2 skills: reading and listening; that's why the last change was done by adding writing reflective journal based on topic. Brown (2010) states that language skills shouldn't be tested separately as they compensate each other. Regarding to Browns' experience the additional writing task will be extremely productive and meaningful. In order to produce a reliable and practical test a holistic rubric (appendix 2) for writing piece was chosen. According to given rubric, students' work is assessed for content, idea development, organization and mechanics.

Instructions for listening mid-term test:

Time: 65 minutes Total score: 40 points

I. Lead-in questions:

Time: 5 minutes:

1. How much news do you listen to everyday, every week and every month?
2. What are the benefits and drawbacks of news in our daily life?
3. What does "no news is good news" mean for you?

II. Listen to National Public Radio's Talk. Margot Adler interviews people about how much news we need. Take notes, pay attention to 5 people's opinion (whose names are given below and one of their ideas introduced) about news and use your notes to answer multiple choice questions on the next page. You will listen to the recording twice. (12 points for taking notes). Time: 12 minutes



1. **Andrew Weil (doctor, author)** –*People should gradually reduce news intake*

2. **Mark Harris (author)** - *Novels are more interesting than news,*

3. **Gabrielle Spiegel (Johns Hopkins University)** - *Fantasy and humor are more important than the news,* _____

4. **John Sommerville (author, history professor)** - *It's better not to read the news every day,* _____

5. **Tupton Shudrun (Buddhist nun, teacher)** - *The news focuses too much on problems and doesn't offer solutions,*



III. Answer to questions with the help of your notes. (8 points) Time: 8 minutes

14. Mark Harris wrote an essay in the *New York Times* in the early 1970s about newspapers. Since then, he has _____.

- d. changed his opinion completely**
- e. adjusted his opinion a little**
- f. maintained the same opinion**

15. In Harris's opinion, novels get readers to focus on interesting people, such as _____.

- d. sports figures who don't win**
- e. teachers at small universities**
- f. historical figures who are little known**



16.	The period in history that Gabrielle Spiegel studies is the ____centuries.
d.	13 th and 14 th
e.	15 th and 16 th
f.	19 th and 20 th
17.	Spiegel doesn't read newspapers because_____.
d.	they are repetitive
e.	she doesn't have enough time
f.	she wants to spend time with her children
18.	In Sommerville's opinion, daily news_____.
d.	gives too much information
e.	keeps people from seeing connections
f.	should focus more on religion
19.	Sommerville reads the news_____.
d.	every week
e.	every three or four months
f.	almost never
20.	Tuption Shudrun is critical of the media because it_____.
d.	doesn't explain problems well
e.	creates a sense of despair
f.	is difficult to understand
	21. Gabrielle Spiegel thinks that children_____
	d. can't understand much of news they hear
e.	need to spend more time alone
f.	should listen to the news with their parents

IV. Now based on radio talk about news and your own experience write a reflection by answering the following questions.

- What did you learn and find interesting?
- Which of the ideas mentioned in the recording do you agree or disagree with?

Word limit: 150-200

Time: 40 minutes

The modified version of listening test aligns with 5 principles of testing. First of all it is practical since it is easy to check notes and piece of writing with holistic rubric; furthermore, multiple choice questions are easy to score. Moreover, it doesn't cost a lot to print and budget issues are covered by lyceum. The test is reliable as full instructions for tasks were provided and rubrics were selected. Intra-reliability also can be involved if the writings are checked by 2 instructors by using holistic rubrics. The test should measure the knowledge of students that was already covered during the lesson. How to write a reflective journal is supposed to be conducted before the mid-term test according to the syllabus; that is the reason why the test is valid. Although a real version of test material was authentic, it lacked authentic tasks that could be useful in implementing them in real-life situations. However, modified version of test contains a task of writing reflective journal which is useful in further studies of test-takers. This modified test can provide a positive washback since it is scored according to rubrics and students are able to get appropriate feedback for their notes and writings after their tests are checked.

The listening material which is authentic is appropriate for B2 level students as it contains some background noise as well. The note-taking task was redesigned by providing detailed instructions and enriched with a rubric in order to meet learners' level. The tasks are aligned to the proficiency level of test-takers since they should be able to produce a writing piece based on gained knowledge and link it to their own experience.

Although the test was first used some time ago, I decided to administer the original and modified versions to volunteer learner from Westminster lyceum. In taking original version of the test, he faced some challenges in note taking part. Moreover, in matching and multiple choice question sections, he randomly answered to some questions which led to some mistakes. He mentioned that listening material was tough and difficult to concentrate. However, he had a beneficial washback in accomplishing the modified test. The first thing that he appreciated was lead-in questions which helped him to predict the topic. Moreover, in the listening task having names of speakers in the paper and extra ideas that belong to those 5 people were useful for him. He also enjoyed writing reflective journal by connecting his own experience with the gained knowledge from listening extract. By getting a second chance for listening the recording, he was able to produce better work in note taking part. Besides, he didn't make any mistakes in answering to multiple choice questions and he was assured with his answers. Furthermore, he liked the idea of his writing to be checked according to given rubric by 2 instructors as he believes it will be much fair. All in all, he preferred modified version of test as it was much interesting, informative, meaningful and authentic.



Appendix 1

Rubric for assessing notes:

	Wow! 4 points	Yes! 3 points	Yes but...2 points	Almost there... 1 point
Neatness Organization of notes.	Work is exceptionally neat.	Writing is easy to read, neat, no smudges. No signs of doodles. Paper is not crinkled.	Writing is difficult to read. Some smudges and doodles are present. Paper is a little crinkled.	Writing is messy and almost impossible to read. Smudges and doodles may be present and the paper is crinkled and/or ripped.
Points/main ideas Ideas from the discussion.	All points are listed in your notes. This is clear and draws conclusion about the conversation. Notes are easy to read and include elaborate vocabulary.	Some of the five points/ideas discussed are listed in the notes. Notes are easy to read.	Only a few points are listed with little explanation.	Few, if any, of the important information is shown. Sentences are unclear; vague.
Comprehension Understands speakers thoughts and ideas.	Student shows above/high average understanding of the discussion and illustrates it in their notes	Student has above/high average understanding.	Student has basic understanding and can complete with assistance.	Student has little understanding of the material and cannot complete without significant assistance.

Retrieved from <https://www.rcampus.com/rubricshowc.cfm?code=K929B2&sp=yes&>



Appendix 2

Rubric for assessing reflective writing:

Table with 5 columns: Grading Criteria, Excellent 5 points, Acceptable 4 points, Minimal 3 points, Unacceptable 1 point. Rows include Content, Idea Development, Organization, and Mechanics.

Retrieved

from

http://www.rcs.k12.in.us/files/Rubric%20for%20Assessing%20a%20Journal%20Entry.pdf



Reference

1. [Advance Consulting for Education]. (2018, December 26). *The Principles of Testing and Assessment* [MP 4] Retrieved from <https://www.youtube.com/watch?v=VGv-AJVcuW8&list=PLpVVpXGOPYivpZ1dt5Ckp0oBxPwZiym7R&index=3>
2. <https://www.youtube.com/watch?v=VGv-AJVcuW8&list=PLpVVpXGOPYivpZ1dt5Ckp0oBxPwZiym7R&index=3>
3. Bachman, L. F. & Palmer, A.S. (1996). *Language Testing in Practice*. New York: Oxford University Press.
4. Brown, H. D & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). White Plains, NY: Pearson Education
5. Brown, J. D (1996). *Testing in Language Programs*. New Jersey: Prentice Hall Regents
6. Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge England: Cambridge University Press