

UDC 378.147

<https://doi.org/10.33619/2414-2948/55/39>

## ANALYZING TESTING TOOLS ACCORDING TO THE FIVE PRINCIPLES OF LANGUAGE ASSESSMENT

©**Tukhtabaeva Z.**, *Alisher Navoi Tashkent State University of the Uzbek Language and Literature,  
Tashkent, Uzbekistan, zamira.tukhtabaeva@gmail.com*

## АНАЛИЗ ИНСТРУМЕНТОВ ТЕСТИРОВАНИЯ ПО ПЯТИ ПРИНЦИПАМ ОЦЕНКИ ЗНАНИЯ ЯЗЫКА

©**Тухтабаева З. К.**, *Ташкентский государственный университет узбекского языка и  
литературы им. А. Навои, г. Ташкент, Узбекистан, zamira.tukhtabaeva@gmail.com*

*Abstract.* In order to have properties of a good test, assessment tools should meet the special requirements. In the current mini research, the selected test is analyzed and modified according to the five principles of language testing: practicality, reliability, validity, authenticity, and washback.

*Аннотация.* Для создания хорошего теста инструменты оценки должны соответствовать особым требованиям. В текущем кратком исследовании выбранный тест анализируется и модифицируется в соответствии с пятью принципами языкового тестирования: практичность, надежность, валидность, аутентичность и обратный эффект.

*Keywords:* assessment, practicality, reliability, validity, authenticity, washback.

*Ключевые слова:* оценивание, практичность, надежность, достоверность, аутентичность, обратный эффект.

Assessment is one of the influential issues in language education, that is, its impact on student's learning is powerful. It can either encourage the student to learn the target language further or demotivate him/her to the level that he/she may not want to acquire it anymore. Therefore, every assessment technique and tool are to be analysed before putting them into practice. This requires in-depth knowledge from teachers about the main principles of language assessment. Only in this case, teachers can evaluate the testing techniques/tools against those principles and implement an effective assessment procedure. In the current article, I am going to present a mini-research which shows the analysis procedure of a testing tool, and the modification process of that tool in order to make it more appropriate and beneficial for the test-takers.

The main principles of language testing are practicality, reliability, validity, authenticity, and washback. They are briefly explained one by one in paragraphs below, and the selected testing tool is analysed against these principles.

As A. Brown stated "practicality refers to the logistics, down-to-earth, administrative issues involved in making, giving and scoring an assessment instrument" [1, p. 16]. And reliable test (reliability) should have similar results when administered on various occasions and/or with different students. In addition, it should give clear directions and have uniform rubric for scoring; should "contain items/tasks that are unambiguous to the test-taker ... The items need to be evenly distributed; distractors need to be well designed" [1, p. 27–29]. As for the validity, it is one of the most important features of a good test, and a valid test:

–Should measure exactly what it proposes to measure.



- Should not measure irrelevant or “contaminating” variables.
- Should rely as much as possible on empirical evidence (performance).
- Should involve performance that samples the test’s criterion (objective); should offer useful, meaningful information about test taker’s ability [1, p. 30].

When it comes to authenticity, this can be achieved by using the words and phrases that are natural as possible; contextualizing instead of using isolated items; including meaningful, relevant, interesting topics; simulating real world tasks.

The last principle washback is defined by K. Bailey as the influence of testing on teaching and learning [2, p. 221]. Hughes points out that backwash can be beneficial or harmful for teaching and learning according to the nature of its influence. Sometimes assessment tools and their results bring changes to the teaching approaches and methods, or to the whole language program. Sometimes the test-takers become motivated to enhance their language skills further even if they have not achieved their desired score. In this case, the test achieves its beneficial washback in teaching and learning a language. But sometimes the test itself can be invalid and unreliable or does not provide any feedback to reflect on. In this case, the test yields harmful backwash hindering both teaching and learning processes [3].

### *Learner profile*

The target learner is an eleven-year-old primary school student who started to learn English from the first grade at School no. 271 in Yunusobod, Tashkent. I taught her class during 2018–2019 school academic year when she was a beginner level (CEFR A1) student. The reason I had chosen her for this research was her growing enthusiasm and effort to learn English. Her hard work and interest towards English always reflected on her performance in the lessons as well as in classroom tests.

### *The description of the chosen test (See Appendix 1 for the test)*

The chosen listening test that the learner took in January 2019 was the part of Olympiad test presented by the Department of Public Education of Yunusobod district (Tashkent) to check 4<sup>th</sup> graders’ listening comprehension in the School Olympiad. The test checked ‘Listening for detail’ sub-skill and contained a recording, an instruction and ten question items to assess the listening comprehension of beginner level students. The question items were based on one type of task which asked participants to choose from two options in the sentences according to the recording and underline the correct answer, and test-takers get two points for each correct answer. The allotted time for administration was 10 minutes including organizational issues: distributing stamped papers, explaining the rules to the participants and administering the test itself.

### *Analysis of the test*

*Practicality.* The selected test has the following features of practicality:

- Easy to develop as it has only one type of task for checking listening comprehension and uses the same sentences in question items as in the input.
- It is easy to administer as long as the technical equipment needed to play the recording is available and in good quality.

The exact answers to the questions make it highly practical to score the test-taker’s work.

*Reliability.* The selected test provides testers with clear directions for scoring, but it does not yield objective result due to the certain flaws concerning with the design of the questions and distractors.

*Validity.* The chosen test measures what it is intended to measure, it does not measure other irrelevant variables, and the content is familiar to the learner. But the face validity is questionable because of some existing good and flawed features in the test. For example, the format is clear and tests with the same task was taken in the previous progress tests, the instruction is short and understandable. Yet, some answers should be spotted immediately one after the another, that is, there is a high possibility to miss the answer to the next question. What is more, the sequence of some question items does not suit with the recording. Therefore, the answers cannot be found in consecutive order which makes the test very confusing for the learner.

*Authenticity.* The question items of the chosen test contextualized but the test itself does not fully replicate real world situation.

*Washback.* Washback effect cannot be analyzed through the test as it does not intend to provide any feedback to the learner except the test result.

Also, the chosen test is free from any bias mentioned above, and it is equally valid for all 4<sup>th</sup> grade students considering their needs, the content and components of the test itself. Besides these, the test is level appropriate for 4<sup>th</sup> graders according to the ESL requirements aligned with CEFR standards.

### *The critique of the listening test*

This particular test has several features of a good test. In the test:

- The content and the wording of the recording were familiar to the student as she learnt this topic and related words in the first quarter of the 4<sup>th</sup> grade.
- The rate of the speech in the recording was suitable for the level of the learner (not too fast).
- The test format was clear, and the instruction was short and understandable.

In addition to abovementioned factors, this type of test had been taken by the learner in class and progress check tests many times which improved face validity of the test by lowering the anxiety of the learner. However, the presented test also has certain flaws which impact on student's result and do not meet the requirements of a good test. They are the followings:

- An example is not provided for the task.
- He answer of some test items can be easily guessed, that is, the distractors are not well designed (e.g. test item no. 7 can be answered according to the verb “watch” even not listening to the recording).
- The wording of the test item is the same as the test stem that may influence on objective evaluation of the learners' listening comprehension.
- The place of some correct answers in the recording are very close (e.g. questions no. 1 and no. 2).
- The order of test items is not compatible with the recording (e.g. question no. 4).
- There is only one task type of testing listening.

After the analysis, I have decided that this test should be redesigned to meet the standards of the good test and to obtain objective result.

### *Modification of the target test*

Teaching foreign languages from early childhood, creating materials and assessment tools for young learners is a recently introduced field in our country. Admittedly, the quality of available resources for teaching foreign languages in public school system needs improvement in large scale. Here, I would like to define the term ‘young language learner’ as it is the main ‘actor’ of the current research. P. McKay states that young language learners are the students who are learning a new language apart from their L1 (first language) “during the first six or seven years of formal

schooling” and their ages range approximately from five to twelve [4, p. 1]. The target student also belongs to this category and has certain characteristics that are peculiar to young learners in terms of background schemata, cognitive, emotional and physical features. Therefore, these features are very important to be considered for the effective assessment of the young learner. To be more specific, the weight of the assessment task must be suitable with her age-related capabilities and the input used in the task should be taken from familiar content and genre for her. In addition, to grab her attention and prolong her concentration span the test should also include colorful pictures and different kinds of tasks, not only plain format and questions that ask to choose from two options.

Besides abovementioned factors, the assessment procedure in L2 is not the same as the procedures implemented in native language of the learner to assess her knowledge of other subjects: mother tongue (Uzbek), math, nature studies and so on. It is true that listening process is similar in both L1 and a foreign language: recognizing sounds and words, decoding their meaning, comprehending the utterance using background knowledge. What differs in this situation is “foreign language listener has a restricted knowledge of the language, and that they might be influenced by transfer from their first language” [5]. When the learner of L2 (foreign language) is listening, he/she may only understand isolated words or phrases if the spoken message or its rate is not compatible with the level of the learner. Of course, this gap can be observed in first language too, but its impact on comprehension is more significant in L2 listening as the learners often lack the needed textual schemata and linguistic abilities of the new language to compensate for occurred gap.

Taking into account all specific features related to the target student and assessment procedures the following reasonable suggestions are recommended for modification of the original listening test:

The type of tasks should be diversified to gain more objective results and increase reliability of the test itself. But while choosing task types, the test developer must consider the familiarity of the test-taker with presented tasks in order not to lose validity of the test and lower the test anxiety of the student.

One example should be provided for each type of task in order to assure the test taker what is expected from her in the particular task type.

The sequence of questions should be compatible with flow of the recording.

The place of correct answers in the recording should not be too close in order to allow some time for the test-taker to spot the answer and respond to the question.

Distractors in the options should be selected reasonably according to the attractiveness as well as the level of the test-taker.

The question stems and options should be paraphrased where appropriate to obtain objectivity in assessing the target skill of the learner. This listening task shouldn't only involve recognizing the sounds or single words, but it should deal with overall comprehension of spoken message.

Pictures should be included to grab the young test-taker's attention and avoid boredom and fatigue during the test.

Considering all suggestions above, the selected assessment tool is redesigned to meet the requirements of a good test (See Appendix 2).

#### *Administration of the modified test*

When the modified test (See Appendix 2 for the test) was administered, the student scored 14 out of 20 which means she answered incorrectly to three questions compared to two mistakes in the original test. Then the answers in both tests were analyzed thoroughly to identify the reason for lower performance. The following tables show the answers to questions of both tests. In the tables, a “+” denotes a correct response, while a “-” represents a mistake:

Table 1.

ORIGINAL TEST RESULT

Test №	Que №1	Que №2	Que №3	Que №4	Que №5	Que №6	Que №7	Que №8	Que №9	Que №10	Correct answers / score
Answers	+	+	+	+	+	+	+	–	+	–	8 /16

Table 2.

MODIFIED TEST RESULT

Test №	Que №1	Que №2	Que №3	Que №4	Que №5	Que №6	Que №7	Que №8	Que №9	Que №10	Correct answer / score
Answers	+	+	+	–	+	+	–	–	+	+	7/14

According to the first table, the student answered correctly even to flawed question items, which in reality cannot be answered appropriately. For example, Question no. 4 asks the learner if the girl has her dinner at home or at the school canteen. But the correct answer to this question comes later in the recording. Here, the test-taker either guessed the right answer, or chose it reasonably as dinner is always eaten at home as the latest meal in the evening.

In the modified version, the student made mistakes in Questions no. 4, 7 and 8. The answer to the question no. 4 was marked as incorrect answer due to the fact that the student underlined both answers. Here the student might have confused the days of the week. In the 7<sup>th</sup> and 8<sup>th</sup> test items, paraphrased question stems made it difficult for learner to find the right answers.

All things considered; the following conclusion can be drawn from the modified test result:

- The diverse task types were challenged the learner leaving little room for guessing.
- The reasonable distractors showed the objective performance of the learner.

Paraphrased questions and options checked the real comprehension of the learner revealing her weak points in listening comprehension.

After the learner became aware of her result, she was provided with necessary feedback and analysis of her result for future consideration in the learning process.

During the process of this study, I have gained invaluable expertise in terms of selecting appropriate testing tools and evaluate them against the five principles of language assessment. Now I have clear understanding that the available tests should be adapted if they do not meet the requirements of a good test and the needs of the target learner. What is more, now these five key principles of language assessment have become more comprehensible to me as a result of the practice I had in the research.

*Appendix 1. Original test*

Listening Test for the 4<sup>th</sup> grade

Listen and underline the right answer.

Becky goes to school in the morning/afternoon.

She goes to school on foot/by bus.

She does her homework at home/in the library.

She has dinner at home/at the school canteen.

She goes to the library after/before lunch.

Before dinner she usually plays with her cat/dog Blacky.

She watches her favourite music/cartoon.

She always/never goes to bed late.  
On Sunday and Saturday, she goes to school/stays at home.  
Her favourite day is Saturday/Sunday.

*Appendix 2. The modified version of the test*

Listening test for the 4<sup>th</sup> grade

A. Listen and underline the right answer in questions 1–4.

There is given one example.

e.g. The name of the girl is Katy / Becky.

1. She lives in London/New York.

2. She goes to school by car/bus.

3. She usually has five/six lessons

4. The girl goes to music club / tennis club on Tuesdays and Fridays.

B. Listen and circle the correct answer in questions 5–7. There is given one example.

e.g. She usually has lunch at the school canteen.



a) never

b) usually

c) sometimes

5. She goes to \_\_\_\_\_ after lunch.



a) school library

b) music club

c) tennis club

6. Before dinner she usually plays with her cat \_\_\_\_\_



a) Kitty

b) Lucky

c) Blacky

7. She always goes to bed \_\_\_\_\_



a) early

b) late

c) at 11 o'clock

C. Listen and write T for True, F for False sentences. There is given one example.

e.g. She never goes to school on Saturday and Sunday.   T  

8. She doesn't go to the supermarket with her father.           

9. She gets up late on Sunday.           

10. On Sunday she visits her granddad.           

*References:*

1. Brown, A. (2014). Pronunciation and phonetics: A practical guide for English language teachers. Routledge.
2. 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>
3. Hughes, A. (2003). Testing for language teachers. Ernst Klett Sprachen.
4. McKay, P. (2006). Assessing young language learners. Ernst Klett Sprachen.
5. Buck, G. (2001). Assessing listening. Cambridge University Press.

*Список литературы:*

1. Brown A. Pronunciation and phonetics: A practical guide for English language teachers. Routledge, 2014.



2. Bailey K. M. Working for washback: a review of the washback concept in language testing // Language Testing. 1996. V. 13. №3. 257-279. <https://doi.org/10.1177/026553229601300303>
3. Hughes A. Testing for language teachers. Ernst Klett Sprachen, 2003.
4. McKay P. Assessing young language learners. Ernst Klett Sprachen, 2006.
5. Buck G. Assessing listening. Cambridge University Press, 2001.

*Работа поступила  
в редакцию 28.04.2020 г.*

*Принята к публикации  
01.05.2020 г.*

---

*Ссылка для цитирования:*

Tukhtabaeva Z. Analyzing Testing Tools According to the Five Principles of Language Assessment // Бюллетень науки и практики. 2020. Т. 6. №6. С. 298-304. <https://doi.org/10.33619/2414-2948/55/39>

*Cite as (APA):*

Tukhtabaeva, Z. (2020). Analyzing Testing Tools According to the Five Principles of Language Assessment. *Bulletin of Science and Practice*, 6(6), 298-304. <https://doi.org/10.33619/2414-2948/55/39>