THE IMPORTANCE OF FIVE MAIN PRINCIPLES OF LANGUAGE ASSESSMENT IN DESIGNING THE LANGUAGE TESTS IN UZBEK UNIVERSITIES

Feruza Tulibayevna Jumaniyozova

The Efl/Esl Teacher Of The Uzbekistan State World Languages University Tashkent, Uzbekistan

ABSTRACT

Language testing plays a vital role in the language learning process, enabling teachers to diagnose students' strengths and weaknesses, which place them at their appropriate levels. This article discusses three significant aspects of L2 test design: (1) the importance of five primary principles of language assessment; (2) the disadvantages of using only limited production and selected-response tasks; (3) the importance of using extended tasks in testing students' grammar skills. I researched in the Uzbekistan state world languages university and concluded that the efficient assessment, which follows five fundamental principles of language assessment, facilitates learners' further study and achievement.

KEYWORDS: - Limited production tasks, selected-response tasks, higher-order thinking skills, lower-order thinking skills, practicality, validity, reliability, washback, face validity, authenticity.

INTRODUCTION

Testing and assessing students' knowledge of the target language is as essential as teaching the language since only by evaluating the learners' language skills we can decide what they need to improve or eliminate. Testing English language levels provides valuable data about what the students know and what they should learn in the future. On creating language tests, it is crucial to consider five primary principles of language assessment, such as practicality, reliability, validity, authenticity, and washback. Language instructors should understand the power of tests and their importance in the lives of learners. Low stake tests can motivate students to learn the target language in more depth, whereas high stake tests may open doors for excellent education and desired professions.

Learner's profile

The subject I worked with is a current freshman student at the Uzbekistan State World Languages University (UzSWLU) studying English philology. Denise (the name is disguised), 18 years old, is an only child in a family. Regarding nationality, she is a Tatar who has been raised in a traditional Tatar family with strong cultural and spiritual values. Denise is monolingually Russian, Tatar, and Uzbek fluently, using all three languages in and outside the home and in educational domains. She also speaks good English because of being exposed to learning it

since she was in fifth grade at school and taking extra tutorial courses in several educational centers. She had been learning the English language for six years before she entered the university in 2019. Denise is friendly speaking to her all friends politely, open-minded-always considering others' feelings and opinions, and diligent-being thorough in doing the class and home assignments.

The primary reason I chose this student is that I have known this learner since 2017, and it was the time when I was working a full-time job at the academic lyceum under World Languages University, so I was her English teacher during the curriculum 2017/2018. She was very hard-working and studied hard to enter university. Among her classmates, Denise was outstanding with her writing and speaking abilities, and she was always motivated to learn languages. Denise passed the IELTS exam on the 1st of June in 2019, scoring seven in both writing and speaking. Her overall score is 6.5 representing B2 level 'competent user.'

For the Test Modification Project, I chose the ultimate test from the module of Grammar, which was given to the group mates of the subject as an ultimate grammar test at the end of the first semester. The reason I chose the test from Grammar is that from the second semester, I have been their grammar instructor, and it is exciting for me to know whether my students have a positive or a negative washback and face validity. Analyzing their opinions helps me create a final test that can meet their demands at the end of the second semester. I modified the test, which Denise solved in the exam since I found some drawbacks and deficiencies of authentic materials.

According to Denise, the new test is more exciting and more accessible because it includes many real-life tasks like reading and writing. In the interview with Denise, she says, "I found the new test more comfortable, and by doing the test, I realized that doing grammar tasks should not be only separated sentences. It can also be reading, writing, and speaking tasks since we are learning grammar for not doing only exercises but for being able to use different complex grammatical structures in real life." She also mentioned that grammar tasks in separate sentences are challenging to understand and easy to forget quickly. However, producing a spoken or written language using grammar rules that the students learned during the course is more beneficial to remember the rules for a long-term period.

"Another aspect that I like about the new test is task 3 in which the myth is given about the Greek god of the sea, Poseidon," says Denise, "This was the most natural part for me to do since I love reading myths. However, in the old test, I was a little confused to find the correct answers as there were only sentences, not the entire text to comprehend." She also adds, "I liked task 5 because I think we should learn how to use complex sentences and grammar rules in writing as we often have to pass international exams like IELTS or TOEFL. So I consider structuring correct sentences is vital in essay writing in these exams."

Being one of the top students in the class, Denise has bright ideas about what she needs and what is unnecessary for her in learning a language. Besides what I said above, I would say that Denise is a visual and kinesthetic learner. She enjoys being involved in different activities in class, and she finds it enjoyable to use the language in practice rather than learn by heart the theories about it.

Critique of an existing test

The test I chose for analysis and modification is the final test of the grammar module at the end of the first semester at the Uzbekistan State World Languages University. The test was

prepared in ten options for students, and I analyzed variant three as it was the test that the student I worked with did in the exam. The test has four parts created using the traditional method, and the students were supposed to take 0.25 points for each task. Regular tasks like Limited Production and Selected Response were used in the test where the test-takers could find answers by guessing or memorizing some distinct phrases. Dialogue Completion and Gap Filling tasks were made using straightforward sentences.

Language teaching is one of the challenging tasks since it requires time and effort from the instructor. The instructor should be careful in teaching and testing learners because the way of teaching or testing may work as a motivation or discouragement for the students. Considering five fundamental principles of assessment, such as practicality, validity, reliability, authenticity, and washback, is the central part of designing language tests. In this test, it can be seen that some parts of the test do not address these fundamental principles. For example, according to the syllabus of the course, students should gain the target language at the C1 level according to European standards. The students at the university are supposed to study grammar for two years and considering that it is their first year, they should be learning at least B2 level grammar. The test was not designed based on the B2 level grammar specifications. There are neither any grammatical structures in the given sentences nor any high-level vocabulary. Therefore, it is easy to believe that the test lacks content validity. To reach this end, teachers should attempt to design the test carefully.

On doing the test, if the test-takers and instructor can measure what the test-takers learned during the course, then it can be considered that the test is adequately valid. Content validity is closely connected with Face validity since Face validity is the students' assessment of the test. The test-takers study at the university and many of them have got B2 certificates given from The State Testing Center and 6-6.5 IELTS bands presented by the British Council. The tasks given in the final test are not challenging enough to give them further motivation to study the language in more depth, and this may affect the consequential validity of the test.

As for practicality, this test is efficient as it involves only Limited Production and Selected Response tasks; it does not take much time in both designing and scoring. The test requires one-point answers, which are simple to respond to and check at the same time. On doing these kinds of tasks, students do not use their higher-order thinking skills, and they do not have to produce anything new; they use only given words, or they choose words or structures by memorizing or guessing. The test does not require one-on-one proctoring; all the students can take the test in one lecture room. Testing stays within the budgetary limits since the only item that should be distributed to the students is the printed papers. According to Brown (2010), practical test stays within the budgetary limits, considers the time and effort in designing and scoring, and can be completed by the test-takers within the time constraints (p.26).

Fixed-response tasks have helpful as well as disadvantageous. One superiority of this test is that it requires only one answer in each gap, and it makes the test reliable. First, limited production tasks increase Student Reliability because students do not feel pressure during the test because of the ready options they are given. Second, aim tests also enhance Rater Reliability and Test reliability as these kinds of tests do not demand rubrics, which makes the test reliable in any situation. Using fixed-response tasks is sometimes even beneficial in checking students' fundamental knowledge on using grammar rules or their awareness of some grammatical terminology (Efeator, 2014, p. 208). Subjective

tests, on the contrary, require a rubric for the extended tasks, and it may affect the Rater Reliability.

However, these kinds of tasks require students to produce a limited response by which it is difficult to measure a student's real gained knowledge on the topic. Students cannot show higher-order thinking skills such as those given in Bloom's Taxonomy since it is difficult to perform the analysis, synthesis, and evaluation with fixed-response items. Applying only these kinds of tasks on designing the test may affect the authenticity of the test as they limit the usage of grammar in real-life situations, and the test presented at the university includes only Limited Production and selected-response tasks, which damage the authenticity of the test.

The subject I worked with got an overall 80% from the test; however, tasks with separate sentences made the test ambiguous. According to the students who took the test, the test was ambiguous, and it was easy to find the answer by guessing. The options given in the test made the test convenient to do, and it was unnecessary to produce anything additionally. We can see that the test acts as a positive and negative washback simultaneously. The beneficial washback of the test is its simplicity; students found the test accessible since it requires limited answers. However, the harmful washback is that it may discourage the students from studying further, and as a result, students may not take it seriously about learning grammar in more depth. "Negative washback is considered as the negative influence of tests in the process of language teaching and learning" (Beikmahdavi, 2016, p. 132).

METHODOLOGY

Modification of the chosen assessment

We are living in an era that one cannot justify the separate testing of grammar. In the past, it was fashionable to check students' grammatical knowledge using direct testing; however, now linguists and language teachers understand it is language skills that can be tested, not the abilities (Hughes, 2016, p. 172). Testing grammar is technically known as a form-focused assessment, and the knowledge of grammar lies at the core of language use. Nowadays, in many norm-referenced tests, the knowledge of grammar is checked through reading, writing, listening, and speaking skills. Because grammatical competence is integral to language use, it sometimes helps to check students' knowledge of grammar using direct and indirect assessment.

The first main modification made in this test is that it is changed from mechanical tasks to more authentic real-world tasks. In teaching and testing a language, using authentic materials is essential since naturally testing a language fosters students' autonomy. Choosing real-life materials on the test design increases students' interest in doing the tasks, and they do the tests not only using theories they learned but also practical situations they experienced. As Dan Valeen states that Language Assessment is "a variety of formal and informal procedures that are now used to find out what foreign language learners know and can do in a language continuously" (Dan Valeen, 1999, p.2). Authenticity is a significant factor in students' continuous performance, and it also helps testees in their future performance in their real-life situations.

The second substantial modification is that tasks in this test are changed from decontextualized texts to contextualized ones. For instance, task 1 in the old test was given in separate sentences even though it was a dialogue completion task, which makes it unclear for the test-takers. The student I worked with has a negative washback towards the first task, so I have changed it into contextualized text so that it can be

understandable for the test-takers. In this way, not only the students' knowledge of the use of grammatical forms but also their awareness of the grammatical and pragmatic meaning of those forms can be tested. In the first part of the test, the multiple-choice task is used because Limited Production tasks put the students at ease and avoid unnecessary stress.

The segregated approach has been applied in testing grammar for many years; however, it has been approved that language instructors can teach grammar and test students' grammatical knowledge integrating other language skills. For example, task 3 in the old test was about testing students' awareness of pronouns, and in the new test, this part is integrated with reading about Greek God Poseidon (Poseidon, God of the Sea, n.d.). In this task of the first test, there were given separate sentences taken from a Greek Myth about their God Zeus, and this was the part that the student I worked with had the most mistakes. The student explained that she had little knowledge about Greek Myths, and there were separate sentences that make her feel confused. In the new test, this part is changed into indirect testing in which students do testing through reading. As Douglas states (2015, September 30), "A better way to test is to design an activity that realistically integrates the various language skills. An integrative test draws on a variety of sources. Syntax, vocabulary, "schema," cultural awareness, reading skills, pronunciation, and grammar, are all factors the test maker and test taker need to keep in mind.

The integrative test is generally considered as being a more reliable instrument for measuring language competence." The last modification made in this test is task 5, by which students' productive grammatical writing ability is tested. In this task, students are asked to write a short story on a free topic and use five adjectives, five adverbs, and five-time expressions in their story, so it allows instructors to check students' ability to use correct grammatical structures in production. This task encourages students to use adjectives, adverbs, and time expressions effectively by transferring their grammar skills into writing. In this way, students learn how to produce grammatically correct sentences in writing, which is a vital skill in written language. Learning grammar is not only memorizing structure and using it, but also learning grammar requires higher-order thinking skills such as synthesis, analysis, and evaluation, and students focus on the grammatical and pragmatic meanings of used structures. As Teimourtash (2017) stated, "... higher-order thinking is the key to the development students are expected to make during/at the end of the educational program."

RESULTS AND DISCUSSION

The subject I worked with took the test on the 28th of February 2020, and achieved the following result, which is 12.5% more than prior result:

Tasks	Task 1. Modified	Task 2. Original	Task 3. Modified	Task 5. Modified task
Total score	10 points	10 points	10 points	20 points
Received scores	10 points	10 points	8 points	18 points
Percentage	100%	100%	80%	**90%**
Overall score		92,5%		

Table. 1. The test-taker's result.

In the interview, the student I worked with, Denise, said that the second test was more natural than the first one. She said that she especially enjoyed reading and writing parts as she prefers to produce new sentences rather than put some given words in the gaps of structures. We can see that the learner is in the correct way to learn the language, and she has already understood the importance of authentic materials and likes to challenge herself in some real-life situations, knowing the benefits of that for enhancing language skills.

Because of the test modification, the quality of the test has improved in a way that allowed the students to perform the all-new knowledge they have gained. For instance, the changes made in the experiment led to the following results:

1. The new test is reliable in a way that the test-takers feel little pressure during the trial and feel psychologically comfortable because the test starts with the selected response tasks and follows by extended tasks.

2. Grammar tasks include authentic tasks like reading and writing, and it allowed them to increase the productivity of performance.

3. Adding a writing task and reading task and because of some other changes, the content of the test has improved, making some enhancements on the face validity.

4. The rubric prepared for the extended task also contributed to the enrichment of the practicality of the assessment. Any rater can easily follow the criteria of the rubric to mark the students' essay, and it does not influence the rater's reliability.

CONCLUSION

To conclude, I would like to say that as language teachers, we should not try to find a simple way

to test our students, but we should create a vital way to check our students' knowledge. As Hughes stated (1989), "There is a tendency to test what it is easiest to test rather than what is most important to test" (p. 44). This may be because examiners consider that subjective scoring may influence the reliability of the test. By only checking students' language production skills, we can achieve an expected result, and by doing these kinds of tests, learners may know their strengths and weaknesses. This can help them work on their weak points. In this way, the assessment of language turns into an assessment for language. Another essential aspect is that by following the five fundamental principles that Brown mentioned in his book called 'Language Assessment: Principles and Classroom Practices,' language instructors can create an effective assessment. According to Brown (2010), designing language tests following these five fundamental principles makes the assessment effective. Developing efficient assessment can help to take positive washback and encourage students to study further. If the test has precise specifications, learners will have a clear picture of what they should achieve. The examiners should ensure the specifications of the test are understood by test-takers. During the exam, instructors should assist if it is necessary.

REFERENCE

1. Brown, H. D., & Abeywickrama, P. (2010). Language assessment, Principles and Classroom Practices (2nd edition). Pearson Education.

2. Beikmahdavi, N. (2016). Washback in Language Testing: Review of Related Literature First. Vol. 1, Issue 4, pp. 130-136. Retrieved from: http://ijmltl.com/fulltext/paper-17092016170302.pdf

3. Douglas, M. (2015, September 30). Multibriefs: Exclusive. Integrative tests: A better way to assess learners. Retrieved from: http://exclusive.multibriefs.com/content/integrative-tests-a-better-way-to-assess-english-learners/education

4. Dooley, J. & Evans, V. (1999). Grammarway 4 with answers. Adjective-Adverb-Comparisons. Pp. 42-53. Express Publishing.

5. Efeotor, V. (2014). Fair, Reliable, Valid: Developing a Grammar Test Utilizing the Four Building Blocks. (Pp. 203-225), Retrieved from: file:///C:/Users/AsqarFeruza/Downloads/FourBuildingBlocks.pdf

6. Hughes, A. (1989). Testing for Language Teachers. Achieving Beneficial Washback. Pp. 44-47. Cambridge University Press.

7. Hughes, A. (2016). Testing for Language Teachers. Testing Grammar and Vocabulary. Pp. 172-185. DOI: http://dx.doi.org/10.1017/CBO9780511732980.014 Cambridge University Press.

8. Poseidon, God of the Sea, (n.d.). Retrieved from: http://www.talesbeyondbelief.com/myth-stories/poseidon-the-god-of-the-sea.htm

9. Teimourtash, M., YazdaniMoghaddam, M. (2017). On the Plausibility of Bloom's Higher Order Thinking Strategies on Learner Autonomy: The Paradigm Shift. Asian. J. Second. Foreign. Lang. Educ. 2, 14. https://doi.org/10.1186/s40862-017-0037-8

10. Van Daleen, M. (1999). Test usefulness in alternative assessment. Dialog in language instruction. Pp. 1-26.

11. Shomoossi, Nematullah & Tavakkoli, Mansoor. (2009). Authenticity and Authentication in language Testing. Modern Journal of Applied Linguistics. 2. 1-26.