



Journal Website:  
<https://theusajournals.com/index.php/ijll>

Copyright: Original  
content from this work  
may be used under the  
terms of the creative  
commons attributes  
4.0 licence.

## ENHANCING VALIDITY AND RELIABILITY IN SPEAKING TESTS

Submission Date: March 14, 2024, Accepted Date: February 19, 2024,

Published Date: March 24, 2024

Crossref doi: <https://doi.org/10.37547/ijll/Volume04Issue03-07>

Djumayeva Guli Tulkinovna

Kimyo International University In Tashkent, Tashkent, Uzbekistan

### ABSTRACT

This article highlights the importance of validity and reliability in creating speaking tests, particularly for the Russian National Exam Practice Exam Papers. These tests provide a comprehensive assessment of language skills, encouraging meaningful language use. However, improvements include clearer instructions, authentic materials, and revised test procedures for easier administration.

### KEYWORDS

Language acquisition, assess, speaking, test, learners, exam, student, examiners, validity, language use, authentic materials.

### INTRODUCTION

#### Learner Assessment Profile

First of all, it should be noted that there are many characteristics that make the learner's language acquisition experience unique. Though many share similar characteristics, no two learners are identical (Brown, 2010). The learner who is a third year student of Medical College lives in Sergeli District, Tashkent. She is 18 years old and has been learning English for 5 years. The participant studied English while attending public school. As most of the classes were conducted

in Grammar-Translation method she decided to attend extra English courses in order to improve her general language skills. Her goal is to pass skill-based test and enter the World Languages University. Also she believes that learning English will be useful for traveling and communication with foreigners. However, in this situation she is highly motivated to learn the language. Moreover, her motivation is more intrinsic as opposed to extrinsic. There are no outside factors forcing her to learn the language. She could have learned only grammar and vocabulary in order to

pass the standardized entrance exam for the university, but she has chosen to have skill-based test and work on her integrated skills such as speaking, reading, listening and writing. When considering her learning preferences, it is worth mentioning that she prefers visual and auditory style of learning. She remembers fast when the learning materials are shown through pictures or videos. Moreover, she is good at participating in pair work. According to Shaughnessy (1998), an individual's learning style is a way to concentrate on new and difficult academic knowledge or skill, internalize and remember processes. Being an auditory and visual learner she shows quite good results in reading and listening tasks. For instance, in our last summative assessment she scored 24/ 30 in listening and 25/30 points in reading tests. It shows that she has improved her reading and listening skills quite well.

I have been teaching Pre-Intermediate and Intermediate level students for 7 months in a private education center called "Repetitor Servis". Being located in Tashkent, Sergeli district this center has high reputation and offers various English courses including Skill-based and IELTS together with other subjects. I teach Skill-based (A1-C2) courses and these are based on Common European Framework of Reference for Languages. In fact, the aim of this program is to help applicants achieve B1 and B2 levels, improve their language proficiency and enter the universities of their dreams.

Being an intermediate level student (B1) now, our participant took a placement test at the beginning of our course in order to place her into a certain level of the language. According to Brown(2010), " the ultimate objective of a placement test is to correctly place a student into a course or level, a very useful

secondary benefit is diagnostic information on a student's performance, which in turn gives teachers a head start on assessing their students' abilities". The placement test included various tasks such as multiple-choice, gap-filling, matching, true/false, error correction and they were focused mostly on analyzing the student's reading, listening, grammar and lexical competence. Moreover, she was given a task of writing an informal letter and in order to check her speaking skill and she was asked several various questions on familiar topics such as study, hometown, hobbies, family or friends. The results of the placement test showed that the student's level was A 2. This could be explained by the fact that, in the listening task she could generally identify the topic of discussion which was conducted slowly and clearly. In terms of speaking she could construct phrases on familiar topics with sufficient ease to handle short exchanges, despite very noticeable hesitation and false starts. She used sufficient vocabulary for the expression of basic communicative needs and for coping with simple survival needs which are common for A 2 level. After finding out her level I recommended her to join to a certain group where she could work on her weaknesses, strengthen her knowledge and reach B1 level. This test really helped and always helps me in my teaching, because with the help of this I select and set the tasks which are not too easy or difficult but appropriate for students' level. This year my student is going to take a skill-based placement test based on CEFR to apply for the university.

### Critique of an existing test

Practice Exam Papers for the Russian National Exam which is produced by Olga Afanasyeva, Virginia Evans and Victoria Kopylova in 2007 is wide-spread in Uzbekistan too. In fact, the tests in this book are

designed to assess the listening, reading, writing, grammar and lexical competence and speaking skills of students learning English as a foreign language. The tests aim is to measure learners' general language proficiency accurately and fairly, present a positive impression of international tests, prepare learners for national exam. This language test critique will focus on the speaking test of Practice Exam Papers. It should be noted that, the purpose of speaking test is to identify learners' ability of expressing themselves, asking and understanding questions and making appropriate responses, talking freely on matters of personal interest. The students' speaking ability in this test is measured by the help of two tasks: talking on a certain given topic (monologue) and interacting with a partner on a given situation, discussing and giving some suggestions. Specifically, in the first task the student is asked 4 questions about his/her experience being at the dentist and in the second task the candidate is asked to plan a trip with his friend and discuss types of travel packages (Amazon Adventure Jungle tour, 6 countries in 2 weeks European tour, beach resort and ski resort), then choose one of them . Overall, the speaking test takes 10 minutes for each candidate.

According to Gronlund (1998), validity is the extent to which inferences made from assessment results are appropriate, meaningful, and useful. One argument for the validity of this speaking test is that it is a direct test ( testing speaking by speaking). Hughes (2004) suggests that direct testing improves the validity of the test since it promotes authentic tasks. On the other hand, the second task about planning a trip quite can be challenging for some students who have never experienced any of the travel packages which are mentioned above. This could be explained by the fact that, not all students experienced all types of travel

packages and they do not have enough information to discuss them and this could call into question the validity of the test. As a result, students from low-income backgrounds who have never travelled would also fail the test or perform badly without having enough information about the subject. In terms of reliability, the test is conducted by one oral examiner and one candidate which threatens its reliability. Firstly, it is not easy to guarantee the quality and objectivity and consistency of oral examiners although perfect consistency is not to be expected in the performance in an interview. Secondly, it is difficult for one oral examiner to conduct the test and keep the track of the candidate's performance at the same time. However, in order to reduce the student's anxiety and deliver the test in a non-threatening way only one examiner is used for the interview. Finally, some students have to wait for a longer time than others to be interviewed. In this case, fatigue can become an issue for the last individual student and threaten the reliability the assessment.

This speaking test has both merits and demerits. In terms of its strengths, this is a test of real language skills and language use is tested in meaningful and realistic way. Also the candidates can be asked these types of questions in their daily life too and this serves for the tests' authenticity. Another advantage of this test is that the topic of the first task is valid as most of the candidates experienced going to the dentist they can freely talk about it. Moreover, for this task only one examiner is required and this helps the student to feel more comfortable and reduce his anxiety.

While on the contrary, one of its drawbacks is that the second task does not coincide with the framework of an effective interview. According to Michael Canale (1984), "the third phase of the interview need not to

be a separate stage entirely but might be a set of questions that are interspersed into the previous stage”. But in the existing test the topic of the task is totally different compared to the previous stage. Moreover, this task is student-related unreliable. As I mentioned above it includes topics which are not familiar to the student and it can decrease the reliability of the test. Finally, in the second task the student is asked to interact with the examiner and discuss the topic. But it is difficult for the examiner to conduct the test, assess and be the student’s partner.

### Modified version of the chosen test

In order to make the existing speaking test inter-rater reliable I would suggest maximizing the number of the examiners up to 2. Hadley and Mort (1999) note that inter-rater reliability measures the consistency between different examiners. They describe it as the degree of correlation between two or more examiners, with the goal of determining whether they are using the same set of criteria when testing the oral proficiency of their learners. It worth mentioning that, the aim of this is not to achieve exactly the same results between the examiners, but to get similar results. Another suggestion is to change the individual interview into paired assessment. In fact, paired oral

assessments have gained increasing popularity as a method of assessing speaking skills. Moreover, several advantages have been associated with this method, including practicality and authenticity (Taylor, 2003). Another benefit for this suggestion is that my learner is quite good at pair discussions and this can lead to her good performance in this speaking test.

Taking into consideration the needs and preferences of the student I divided the test into 3 parts. In the first phase two students are asked some personal questions about the topic. This is followed by the next stage in which students are given a situation to discuss and a related picture as a help. Here the students’ turn taking ability is also assessed. The reason why I have changed this task in this way is that, when the student is given a picture for the discussion he/she will have more ideas about what to talk. For the third stage of the speaking both of the candidates are shown the video on the projector or screen and after that they are asked some general questions to discuss. In fact, the theme of the video is also related to the previous stages’ topic.

Before describing the modified version of the test I would like to provide the original version first.



16

**C3**

Give a 2-minute talk on going to the dentist.  
Remember to say:

- if you like going to the dentist(why/why not);
- how many times you've been to the dentist;
- what you think about people who are afraid of going to the dentist;
- if you take good care of your teeth.

You have to talk for 1.5-2 minutes. The teacher will listen until you have finished. Then she/he will ask you some questions.

**C4**

You are planning a trip with a friend. Discuss with her/him which travel package you would be interested in and choose one. Here is a list of possible travel packages:

- Amazon Adventure Jungle tour
- 6 countries in 2 weeks European tour
- beach resort
- ski resort.

You begin the conversation and have to decide on **one travel package only**. The teacher will play the part of your friend. Remember to:

- discuss **all** the travel packages;
- take an **active** part in the conversation and be **polite**;
- **come up** with ideas;
- give good **reasons**;
- find out your **friend's attitudes** and take them into account;
- **invite** your friend to **come up with suggestions**;
- come to an **agreement**.

### MODIFIED VERSION OF THE SPEAKING TEST

Part 1 (2-3 minutes)

Tasks: Identifying oneself, giving information about oneself, talking about interests.

Examiner

Good morning/afternoon/evening,

My name is Guli Djumayeva.

**A:** Now what's your name?

Thank you.

**B:** And what's your name?

Thank you.

**B:** What's your surname?

How do you spell it?

Thank you.

**A:** And, What's your surname?

How do you spell it?

Thank you.

( Select one or more question from the list to ask each student)

- Do you like eating sweets?
- How often do you go to the dentist?
- How do you take care of your teeth?

**OSCAR**  
PUBLISHING SERVICES

(Introduction to Part 2)

Part 2 ( 2-3minutes)

Tasks: Discussing the situation, expressing opinion

Examiner Say to both students:

I'm going to describe a situation to you.

A boy had a terrible toothache but he did not want to go to the dentist. Nevertheless, his mother took him to the tooth doctor .

Talk about the reasons which caused his toothache and how does he feel now.

Ask both students to look at picture and repeat the frame

I'll say that again.

A boy had a terrible toothache but he did not want to go to the dentist. Nevertheless, his mother took him to the tooth doctor. Talk about the reasons which caused his toothache and how does he feel now.

All right? You have 3 minutes for discussion. Talk together.

Part 3 ( 3-4 minutes)

Task: discussing broader topics

Examiner Say to both students:

You are going to watch a video about healthy lifestyle [https://youtu.be/\\_ISifko4kOY](https://youtu.be/_ISifko4kOY)

After watching it you should discuss the questions together taking turns.

You have 4 minutes for the discussion.

(Select one or more questions from the list to ask the students)

- Is it important to have a healthy life style in modern world?
- Why do many people try to have a healthy life style?
- What makes people to have healthy life style?

Thank you.

That's the end of speaking test.



Weir (1990) stated that “in testing communicative language ability we are evaluating samples of performance, in certain specific contexts of use, created under particular test constraints, for what they can tell us about a candidate’s communicative capacity or language ability”. As we know the Common European Framework of Reference for Language: Learning, Teaching, Assessment is designed to bring together the best of a wide range of different schemes for describing language learning levels. This speaking test was designed according to the CEFR descriptors on spoken production. For instance, the 1st section of the test is targeted at the B1 level and it focuses on personal domain requiring the candidate to respond to a series of simple questions on familiar topics such as going to the dentist and etc. (see table A)

Table. A

SUSTAINED MONOLOGUE: DESCRIBING EXPERIENCE		PROSIGN
B1	Can clearly express feelings about something experienced and give reasons to explain those feelings.	
	Can give straightforward descriptions on a variety of familiar subjects within his field of interest.	
	Can reasonably fluently relate a straightforward narrative or description as a linear sequence of points.	
	Can give detailed accounts of experiences, describing feelings and reactions.	
	Can relate details of unpredictable occurrences, e.g. an accident.	

According to the CEFR, sections 2 and 3 are also targeted at the B1 level. The tasks focus on assessing students’ conversational and turn taking skills. (see table B).

Table B

SPOKEN INTERACTION	
OVERALL SPOKEN INTERACTION	
B1	Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thoughts on more abstract, cultural topics such as films, books, music etc.
	Can exploit a wide range of simple language to deal with most situations likely to arise whilst travelling. Can enter unprepared into conversation of familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).

A strong motivational characteristic of this modified test is that there is not pass or fail. It is designed to test what students know instead of what they do not know. While scoring the student’s speaking proficiency I took into consideration her language range, accuracy, fluency, interaction and coherence.



Before piloting the modified version of the speaking test I asked my colleague for some help. In order to increase the reliability of the test I decided to have the same assessment rubric with my colleague and then compare the results. I conducted the modified version of my speaking test with my student and found some differences between the results of the original test and modified version. The first thing which should be mentioned is that, in the original version of the test there were two topically unrelated tasks which confused the candidate and caused anxiety. As a result it affected the assessment of the speaking proficiency of the student. Because being unable of using the ideas and vocabulary which are related to the topic the student performed the test bad. While on the other hand in the modified version of the test all of the 3 stages are coherent. In fact, I used some visual and technical tools such as authentic picture and video [https://youtu.be/\\_ISifko4kOY](https://youtu.be/_ISifko4kOY), which helped my student to get some ideas and answer the questions thinking critically. Also I changed the type of the second task into pair discussion. According to Brown

(2010), “As formal assessment devices, discussions and conversations with and among students are difficult to specify and even more difficult to score. However, as informal techniques to assess learners, they offer a level of authenticity and spontaneity that other assessment techniques may not provide”. While observing students’ discussion I took into account their eye contact, turn taking, body language, clarifying, questioning and politeness. It worth mentioning that the candidate felt herself very comfortable with her partner while interacting and this lead to good performance. But in the initial version it was challenging to analyze the candidate’s these skills as the task required interacting with the examiner. In order indicate the proficiency level of the student’s speaking skill I used the table for assessing spoken performance based on CEFR (see Table C). It should be noted that, the decision which we made with my colleague in terms of my student’s speaking proficiency were similar and it showed that the candidate’s oral proficiency level is B1.

Table C

	RANGE	ACCURACY	FLUENCY	INTERACTION	COHERENCE
B2	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, without much conspicuous searching for words, using some complex sentence forms to do so.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding, and can correct most of his/her mistakes.	Can produce stretches of language with a fairly even tempo, although he/she can be hesitant as he or she searches for patterns and expressions, there are few noticeably long pauses.	Can initiate discourse, take his/her turn when appropriate and end conversation when he / she needs to, though he /she may not always do this elegantly. Can help the discussion along on familiar ground confirming comprehension, inviting others in, etc.	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some “jumpiness” in a long contribution.
B1	Has enough language to get by with sufficient vocabulary to cope as him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Uses reasonably accurately a repertoire of frequently used “routines” and patterns associated with more predictable situations.	Can keep going, comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.	Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2	Uses basic sentence patterns with memorised phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.	Uses some simple structures correctly, but still systematically makes basic mistakes.	Can make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident.	Can answer questions and respond to simple statements. Can indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her own accord.	Can link groups of words with simple connectors like “and,” “but” and “because”.

## CONCLUSION

In conclusion, I have found this assignment very useful to consider the aspects of validity and reliability in creation of the speaking test. It has made me look at testing oral proficiency in a more critical way, and to be more aware about the need for validity and reliability. Tests which are given in Practice Exam Papers for the Russian National Exam offer one of the options for comprehensively assessing language skills of the learners. They have significant strengths in encouraging meaningful language use, giving learners a positive impression of international tests. Although it is a good test for assessing learners' speaking skill, some areas for improvement are raised in this paper: improving the tasks by changing their instructions to more clear, providing the candidates with authentic materials which help them think critically and have more ideas about the topic, revising the speaking test procedures for greater ease and accuracy of administration. This experience equipped me with a better understanding of the complex dynamics of oral testing and will certainly be of good benefit for my future professional development.

## REFERENCES

1. Brown, H. & Abeywickrama, P. (2010). Language assessment, principles and classroom practices. New York, NY: Pearson Education.
2. Gronlund, N. E. (1998). Assessment of student achievement. Boston: Allyn and Bacon
3. Hadley, G & Mort, J. (1999). An Investigation of Interrater Reliability in Oral Testing. Nagoya National College of Technology Journal, 35(2), 45-51.
4. Hughes, A (2004) Testing for Language Teachers. Cambridge University Press.
5. Shaughnessy, M. F. (1998, January). An interview with Rita Dunn about learning styles. Clearing House, (71)3.
6. Taylor, L. (2003). The Cambridge approach to speaking assessment. University of Cambridge Local Examinations Syndicate Research Notes, 2-4.
7. Weir, C. J. (1990). Communicative Language Testing. Prentice Hall international