
THE ROLE OF TEST MAKING IN LEARNING PROCESS

РОЛЬ ТЕСТИРОВАНИЯ В ПРОЦЕССЕ ОБУЧЕНИЯ

TA'LIM JARAYONIDA TEST TUZISHNING O'RNI

Dilbar KHASANOVA

*INTERNATIONAL ISLAMIC ACADEMY OF
UZBEKISTAN*

*EFL teacher of the department "Foreign languages".
d.xasanova@jiau.uz*

Hugh CARTWRIGHT

Hoseo University, Cheonan

Assistant Professor

e-mail: hughpc@gmail.com

Website: www.concepttid.com

DOI: 10.47980/TLOI/2020/3/18

Abstract: Almost always with the word "test" a question comes to mind with four or five possible answers. Students usually think of multiple-choice questions (MCQs) when they are told to get tested. Testing means testing students' knowledge using different types of question forms. The proposed article examines the academic performance test, which was developed by the administration of the Uzbekistan English Teachers Association (UZTEA) for Year 2 students of lyceums and colleges. The purpose of this study is to discuss test constructs and propose a more reliable version of them if the reason for the unsatisfactory results depends on the evaluation criteria. Both formal and final tests were developed by the administration of the Uzbekistan English Teachers Association (UZTEA). There are specific assessment criteria and objectives based on the curriculum, consisting of four types of questions, addressing the respective four competencies, that is, to test the ability of students in the four core skills. In our study, only criteria related to reading skills will be considered. Reading comprehension questions will be reviewed following relevant developments and recommendations from experts and testing requirements. The importance of the Test Modification Project for English teachers is that test scores tend to be directly related to the content of the tests. The differences between the tests involved in the study and their modified versions will be considered by the specificity of the tests, based on certain criteria and five principles of test design.

Keywords: *test, student performance, test authenticity, assessment, intermediate test, modification, error, practicality, principles, validity, skills, testing.*

Аннотация: Почти всегда со словом «тест» на ум приходит вопрос с четырьмя или пятью вариантами ответов. Учащиеся обычно думают о вопросах с несколькими вариантами ответов (MCQ), когда им говорят пройти тестирование. На самом деле, тестирование означает проверку знаний учащихся с помощью различных типов

форм вопросов. В предлагаемой статье рассматривается тест на успеваемость, который был разработан администрацией Ассоциации учителей английского языка Узбекистана (UZTEA) для учащихся 2-х курсов. Цель данного исследования - обсудить тестовые конструкции и предложить более надежную их версию, если причина неудовлетворительных результатов зависит от критериев оценивания. Как формальные, так и итоговые тесты разработаны администрацией Ассоциации учителей английского языка Узбекистана (UZTEA). Существуют определенные критерии оценивания и цели, основанные на учебном плане и состоящие из четырех видов вопросов, касающихся соответствующих четырех компетенций, то есть для проверки способностей учащихся по основным четырем навыкам. В нашем исследовании будут рассматриваться только критерии, касающиеся навыков чтения. Вопросы проверки понимания прочитанного будут проанализированы в соответствии с соответствующими разработками и рекомендациями специалистов в этой области и требованиями к тестированию. Важность проекта модификации тестов для учителей английского языка заключается в том, что результаты тестов, как правило, напрямую зависят от содержания тестов. Различия между привлеченными к исследованию тестами и их модифицированными версиями будут рассматриваться в соответствии со спецификой тестов на основе определенных критериев и пяти принципов составления тестов.

Ключевые слова: *тест, качество работы учащегося, аутентичность теста, оценка, промежуточный тест, модификация, ошибка, практичность, принципы, достоверность, навыки, тестирование.*

Annotatsiya: Deyarli har doim "test" so'zi deganda ko'z oldimizga to'rt yoki besh javobga ega savol gavdalanadi. O'quvchilarga test sinovlari topshirish haqida aytilganda, ular odatda bir nechta javobga ega savollarni tushunishadi. Aslida test deganda har xil savol shakllari orqali o'quvchilar bilimini tekshirishga aytiladi. Ushbu maqolada, O'zbekiston Ingliz tili O'qituvchilari Uyushmasi (UZTEA) ma'muriyati tomonidan litsey va kollejlarining 2-kurs o'quvchilari uchun ishlab chiqilgan oraliq test sinovlari haqida so'z boradi. Ushbu tadqiqotning maqsadi - sinov savollarini o'rganib chiqish va agar ularning test jarayonlarida olingan natijalari sababi baholash printsiplari bilan bog'liq bo'lsa, ularning alternativ variantini taklif qilishdan iborat.

Har ikkala oraliq testlar O'zbekiston Ingliz tili O'qituvchilari Uyushmasi (UZTEA) ma'muriyatining mezonlariga asoslangan testlar sifatida ishlab chiqilgan. Tuzilgan testlar baholashga oson va aniq, shuningdek o'quv rejasiga asoslanganligi, hamda to'rt til kompetensiyasini tekshirishga mo'ljallangan bo'lganligi uchun, bu testni ishonchli va amaliy, testing to'rt prinsipidan ikkitasi rioya qilgan deb aytish bo'ladi. Shuni aytish joizki, testing qolgan ikkita spetsifikatsiyasi, ya'ni ishlash uchun qulaylik hamda autentiklik darajasi test talablaridan, bizning fikrimizcha biroz uzoq. Garchi oraliq test sinov savollari to'rtta qobiliyatni tekshirishga qaratilgan bo'lsa-da, ushbu tadqiqot faqat talabalarga ishlashda qiyinchilik tug'diradigan qismini o'rganishga yo'naltirilgan, ya'ni o'qib tushunishni tekshirish qism savollarini rijoqlantirishga. Maqolada yuqorida qayd

etilgan test bo'limlari savollari tilshunos olimlar qarashlarini e'tiborga olgan va test prinsiplariga muvofiq bo'lgan holda tahlil qilinadi. Ingliz tili o'qituvchilari uchun ushbu test tahlili loyihasining ahamiyati shundaki, savollarni test prinsiplari va talablari asosida tuzilish, bo'lajak yoki kutilayotgan natijalarida muhim o'rin kasb etishini ko'rsaib beradi. Tushunarli bo'lishi uchun tayyor testlar va o'zgartirilgan variantlar o'rtasidagi farqlar quyidagi bosqichma bosqich test spetsifikatsiyalariga muvofiq tahlil qilinadi.

Kalit so'zlar: *Test sinovi, talaba ishining sifatini, sinovning haqiqiylikini, baholash, o'tkazish, oraliq, xato, amaliylik, prinsiplar, ko'nikma, testning foydaliligi, ishonchli test.*

INTRODUCTION

Assessment is a famous and occasionally misunderstood term in present educational practice. Learners are prone to think of testing and assessing as synonymous terms, but they are not. The description of the two concepts will be discussed below.

Assessment is "appraising or estimating the level or magnitude of some attribute of a person" (Mousavi, 2009, p. 36). An ongoing process is required for an assessment, and various types of methodological techniques are demanded as well.

When a question is answered by a student, a comment is offered, or a new word or structure is tried out, a student's performance will be appraised by a teacher. No matter what kind of work it is, ranging from simple sentences to an essay, the written work is assessed by a teacher, or possibly by his peer. Regarding the other comprehension-check tasks about reading and listening questions, a productive performance is required that the teacher observes and then implicitly appraises can be appraised by a teacher. A model teacher should never stop to assess students, whether those assessments are formal or informal.

However, Tests, are an assessment genre and they consist of a set of assessments. Test makers prepare them in administrative procedures to measure and evaluate the learners' performances. The testing time and questions are developed in a curriculum at the very beginning of the learning process. A test is considered as an instrument, that consists of procedures, items, techniques, and different tasks.

To be qualified as a test, firstly, it should adhere to some criteria. If it is a multiple-choice question, there should be a prescribed correct answer for any test. Even the speaking questions are also required to be in a written form for an interviewer, or the writing tasks should be checked through a rubric and a checklist.

Secondly, according to Bachman (1990), "a test must measure", which identifies the quantity of a test-taker's performance. Sometimes tests measure general ability, while exact competencies or objectives are checked. A general ability level can be determined via a multi skill proficiency test; or some grammatical pattern

can be assessed in a quiz, which measures specific expertise.

Instruments or measurement for tests can vary, as a classroom-based test can be evaluated by the educators' comments, whereas the result of large-scale standardized tests are recorded by a total numerical score. To be defined as a test, a technique should specify a form of reporting measurement. As an individual's ability or knowledge is measured through a test, testers should know the test-takers' background and learning experience. Sometimes the quality and validity of tests do not match learners' abilities.

Working as a teacher for eight years tends to allow me an opportunity to formulate a sense of what "teaching" is. This conception does not simply mean conducting the lesson by informing the learners. It also demands a strenuous analysis of what the teacher is doing: ranging from giving information to assessing the assignments properly. The results of assessments show the quality of lessons, which impacts the productivity of work. However, the matter considered here is how the assessments are designed, and how valid and reliable they are. Sometimes results can be unsatisfactory due to assessment reliability, even though the learners' lessons are viable and robustly designed. Therefore, I made a decision to analyze the test results of my learners.

METHOD

Participants

The six college student candidates, two female and four male, were between 17 and 18 years old. All of them were students of Year 2 at college in different groups with the same level, A2, according to CEFR. Their background knowledge is based on school materials and currently, General English is the course being taught. All of them were accepted according to state-mandated assessment (15 Grammar-based tests) and their last result: 78-79% out of 100 % (January 2019)

The reason for choosing these students among the groups was the unchanged nature of their results, 78% and 79% in the 1st week of December 2018, and in the 4th week of January 2019 in turn. (Table 2). Although their motivation, attendance, and active participation remained consistent over a interval of two months.

A description of the educational setting

As for an educational setting, where these learners are studying, which was founded based on Resolution № 272 of Cabinet of Ministry of Republic of Uzbekistan on 1996, July 31 "On the Establishment of Banking College" under the sponsorship of the Banks Association. This setting has prepared specialists in the fields of banking, finance, accounting, and insurance. The students are accepted according to the state mandatory exams in Mathematics, English, and native language,

answering 15 questions for each subject. The correct answer is signed by 2 points and the overall score must be 90 while passing score is changeable. According to the performance of pupils, it may fluctuate. In 2017, the highest score was 68 points out of 90 and 322 candidates were accepted and they had been mastered by 57 qualified teachers. The number of graduates, who applied for jobs in public, private, and commercial banks and various companies made up 79.5, while 21.5 % of them continued in tertiary education.

General English commences teaching from level B1 according to CEFR requirements and continues being taught during the entire six semesters. The English lessons are conducted once a week and after the Resolution №143 of Cabinet of Ministry № "On measures to accelerate equipping classes of foreign languages of educational institutions in the Republic with the modern information-communication equipment, technical means of teaching in 2013-2016". (№143, May 23, 2013), the English language rooms have been provided with education facilities.

Data Collection

The learners are assessed by traditional, formative, and summative tests. As the tests are based on the curriculum, they are considered as criterion-referenced tests. Achievement tests are conducted in the 7th week of every semester. As "the appeal of formative assessments is growing" (Ross, 2005), it is vital to analyze the achievement tests and modify them, if necessary. The satisfactory results of formative tests (higher than 70,1% out of 100%) could lead to improving the required results of summative assessments. According to requirements of State Standards of Uzbekistan, percentages of assessment stand for in following grading scale.

Table 1

Percentages	Marks	
85.1-100	"5"	Excellent
70,1-85	"4"	Good
54.9-70	"3"	Unsatisfactory
0-54	"2"	Fail

Regarding learners' results, there was a marginal change (78% and 79%) in performance:

Table 2

Test date	Type	Rea- ding	Wri- ting	Spea- king	Writing	Over- all
		25 %	25%	25%	25%	100%
De- cem- ber 3, 2018	Achi- eve- ment For- mative	22%	20%	16%	20%	78 % Mark: 4
Jan- uary 28, 2019	Final Sum- mative	22 %	18%	17 %	22 %	79 % Mark: 4

When it comes to implications of Reading tests, the lower percentage of results can affect instructions in the classroom, as the teacher can mainly focus on the weak points and time management can suffer, because of too much time given to the mistakes of the learner. Nevertheless, these progress checking tests, particularly the Reading section, have less test usefulness (Bachman & L Palmer A S, 2005, pp. pp 17-19), due to its less considerable relevant link to the Course and rarely impacts on further test results. The reasons for usefulness and low demand to maintain these Reading section tests and some test specifications will be discussed in the following paragraph.

Description of the last achievement test

Both formative and summative tests are designed as Criterion-referenced tests by Uzbekistan Teachers of English Association (UZTEA) administrations. Hence, there are specific grades and lesson objectives based on Curriculum and consist of all four competence questions; the entire achievement test is valid and practical. However, it often seems to be an unreachable goal to progress, due to the number of lessons (only once per week). To give a clear picture of the achievement test, I would like to focus on the description of the recent one, conducted on January 28, 2019. There are 25 questions each for listening and reading comprehension, which were followed by a 20-minute-writing task and the cue-card question to measure speaking skill. However only the Reading Section will be analyzed. As Table 2 above demonstrates, that section faced a marginal decrease in both tests 20% and 18% respectively.

The table below shows the types, required, and gained scores of Gafurjanova Charos's from the achievement test.

Table 3

Skills	Types of questions		Overall Number of tasks	Overall point	Given Time
Reading	Multiple Choice Questions (MCQ)	True/False/Not given			
Given	6	9	5	5	25 min
Found	1	7	5	8	25 min

Data analysis

The reading part of the assessment was much more inadequate for test principles than other sections. Therefore, only Reading Passages are the focus of discussion. 10 question items out of 25 from the sample of the Reading Test consisting of selected-response

questions Multiple-Choice and other 15 questions are True/False Tasks. All of them will be analyzed according to the principles of language assessment. A reading passage is provided.

Starting descriptions with the criterion of an effective test, the Reading Section tests above were not "made from assessment results are appropriate, meaningful in terms of the purpose of assessment" (Gronlund, 1998, p 226), which means validity of the given test is not high. Moreover, the content validity of the Test is somewhat invalid, because the students learned "How to be an active learner and how to choose materials for project work", while the topics about the USA, California and London Bike Ride emerged in all three parts of the test. In terms of face validity, as Gronlund (E, 1998) mentioned, if the learners consider the test as fair, useful and relevant to their future improvement, such tests are believed to be valid, as in both the B1 certificate and State Mandatory exams the students will be familiar with this multiple choice and True/False tasks. In this context, consider the following question:

Question 7 "What happens when fish is pickled?"

- a. It becomes crisp.
- b. It turns green.
- c. It dissolves into the rice.
- d. It is preserved

"Do the following statements agree with the information given in the text?"

TRUE if the statement agrees with the information (T)

FALSE if the statement contradicts the information (F)

GIVEN if there is no information on this (NG)"

Question 17 "Your Start Time is indicated by the dark color of your body number in this pack."

Regarding practicality, it suits the description of Mousavi (2009), as it is easy to administer, score (one point for a correct answer), and undemanding to interpret the results. Nevertheless, given time (25 minutes) is inadequate to cover all 25 questions, which means the test has adequate inappropriate items to be considered as an impractical approach. Besides that, the test lacks clear instructions for Part 1 and Part 2 with the absence of an example. In Questions 9, 12, 15, and 16, the lines are shown while there are no line numbers in the passage, which means directions are not clear.

Take Question 12, as an example: "The author cites the United States Constitution (lines 23–24) to

- a. praise the liberties afforded by the Bill of Rights.
- b. show that the government valued the contributions of its immigrants.
- c. imply that all American citizens are equal under the law.

d. suggest that it did not protect Chinese immigrants from discrimination.”

The next principle of assessment is reliability, as the scholars Bachman (1990) and J.D. Brown (1998) highlighted, the test should be designed according to the level of difficulty, from easy to complicated. In this chosen test, questions begin from Multiple choice options, from Question 1 to Question 16, while the distractors are well designed, such:

Question 10. “Which of the following best describes the approach of the passage?

- a. theoretical analysis
- b. historical overview
- c. dramatic narrative
- d. personal assessment
- e. description through metaphor”

The passage looks like a narrative and analysis based topic, however, a historical overview is mentioned here. In lines 48-49 there is a justification, "...a testimony to their outstanding achievements and contributions"

A lot of tests fail in terms of authenticity, (Chun, 2006) but the analyzed one is a high likelihood of being an authentic test. As the question items in the Reading section are contextualized and interesting topics were chosen as a component of test while the source was taken from real-world materials. Thus, the test can be authentic. Question 23 can be an example of authenticity: "Refreshments are free to all participants during the ride." The learner has to agree with whether this statement is TRUE/FALSE/ NOT GIVEN"

There is a context above, the topic is captivating and the entire sentence sounds authentic.

As giving washback is not provided, this field needs to be spread among teachers, because either positive or negative, washback can impact the testing. If the washback is supported consequently after each assessment, the learners might have been offered a chance to adequately prepare.

RESULTS

Strengths and weaknesses of the test

Despite these biases, there are some strengths(S) of this assessment as well as contrasting weaknesses(W).

S) As the test requires high Reading comprehension, it can suit both CEFR requirements and curriculum standards, which demand the learner to reach the B1 level. This is described in the benchmarking statements as a student who "can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure" (CEFR) before graduating college. The tests prepare the learners whose target is to pass the exams of "Identifying the level of knowing foreign languages" and get a B1 level certificate, conducted by the State Testing Center.

W) Although the types of question items are beneficial, it must be realized that the “level of difficulty was not a key consideration for the test developer" (Green, 2006, p. 105). The whole test is above the level of the learner.

S) If a student is not capable of passing the level test exam, but wants to continue tertiary education, where 30 Grammar and Vocabulary based questions are designed, this test is an obvious sample to prepare. In such state entrance examinations, there are only multiple-choice format questions with four options, in the format of a traditional assessment. Therefore, focusing on discrete options will be a thorough preparation for the state-mandatory exams. It would not be an unfamiliar process for the applicants.

W) However, according to "guideline for preparing tasks" by Brown (2005) and Genesee and Upshur (1996) Grammar and Vocabulary of the text of Reading section should be valid at the appropriate level. This is clearly an issue for this Reading test. If the unfamiliar Vocabulary and Tenses used in the passages were mastered in previous lessons, it would have suited to the Curriculum requirements.

S) Contextualized tasks can be another strength of the test, due to the given context for which a learner can anticipate the answers in the Reading section, which indicates the predictive type of validity is fulfilled.

Result of the original test Result of the modified version	
Overall: 79% out of 100%	Overall: 81% out of 100%
Reading: 18% out of 25%	Reading: 22% out of 25%
Correct answers: 18 out of 25	Correct answers: 13 out of 15
The proportion of improvement: 4%	

The modified version of the chosen test

Suggestions for improvement

Unless all five test principles are applied to the evaluation of any classroom tests, such tests tend to be inappropriate. For this reason, my suggestions will hopefully be how to cover all requirements.

As the given test's validity and reliability are low enough, my concern of suggestions will focus on the improvement of the lacking specifications. Take content validity as an initial phase, it would be an objective test if the Vocabulary of environment, sea, transportation, and social issues were provided in previous Units and reviewed before giving the test. As Swain (1984) mentioned, the role of the teacher is "not to be 'tricky', but to encourage them and bring out the best in their performance". To achieve this point, firstly the test itself must be at an appropriate level. Even though

the Reading passage remains the same and maintains authenticity, it is important still to sort out the question types and analyze the distractors by keeping the same context. It needs simplifying! It is a criterion-referenced test and should adhere to the curriculum requirements. According to the curriculum, the reading text should coincide with what was already taught. Therefore, as Hugas (1989) mentioned, before selecting the text there are some guidelines which should be obeyed, such as: not repeatedly selecting texts which are already available; to choose the appropriate length of text (no more than 1000 word in this case); give the passages which have discrete information; choose interesting topics, rather than topics which are too academic; make sure whether only reading comprehension is tested or not.

Concerning the reliability of the discussed test, it is also analyzed due to the time. To provide a fresh atmosphere, it would be better either to check productive skills isolated from receptive or to shorten the numbers of questions with options. There are five options, which means the rate of distractors is pretty high.

Implication

Design of a modified version of the test with new instructions.

A proper way to commence giving modified versions is shortening the numbers of passages, questions, and alternatives. That would include two passages, 15 question stems with three (A, B, C) options instead of 25 items with five distractors. Part I, due to the topic, would be given to the students, because information about sushi history tends to be both beneficial and helpful. As for Part III, as the topic (instructions of bike riding) is authentic and interesting for the youngsters, it may also remain as the topic to check comprehension. Both the reading passages will be given to applying, after introducing the Vocabulary of Asian food and Transportation.

The questions will be by seven and eight for Part I and Part II (Part III is accepted as Part II), respectively and the overall sketch will be as following.

Sentence completion – Questions from 1 to 4

True/False/ Not Given – Questions from 5 to 7

Matching the heading – Questions from 8 to 12

Multiple choice task – Questions from 13 to 15

Test administration and analysis of outcomes

Reading Passage I

You should spend around 15 minutes for the Questions 1-7 based on Reading Passage below. Please choose no more TWO words from the text, after looking at the example.

Example: Sushi was used to be thought as unpleasant and _____ food before 2970 in America.

Answer: exotic

Questions 1-5

The _____ of making food was altered after refusing traditional diets.

Sushi began to be served in _____ of educational settings.

Except its convenience, sushi is considered as _____ food.

During _____ rice and salt contain the cleaned fish.

Questions 5-7

“Do the following statements agree with the information given in the text?”

TRUE if the statement agrees with the information (T)

FALSE if the statement contradicts the information (F)

NOT GIVEN if there is no information on this (NG)”

Please look at the given example before starting to identify the statements' similarity, contradiction, and absence of information.

Example: Fermented products were stopped to produce in 1824 in Yohei Hanaya. (TRUE)

The vinegar might be kept fermentation process

Raw shrimp is the choicest sushi so far.

Colorful eggs are added to impact the cost of a meal.

Reading Passage II

You have to spend 15 minutes for the Questions 8-15

Please read the passages from A to E and find the most suitable HEADINGS from the list for Questions 8 to 12. Study the given example before beginning.

List of headings:

I. Follow the instructions

II. In case of breakdown

III. Start

IV. if you have to drop out

V. Refreshment stops

VI. ride carefully

Example: Paragraph A: III

8. Paragraph B _____

9. Paragraph C _____

10. Paragraph D _____

11. Paragraph E _____

12. Paragraph F _____

Questions 13-15

Please choose the most suitable answer out of three options. Study the example.

Example: What kind of obstacles can be emerged in the opposite direction?

Other cyclists B. Passengers C. Vehicles

Answer: C

13. Which part of the body is injured mostly during cycling?

A. a hand B. a head C. a leg

14. Which organization usually prepares the drinks?

A. The British Heart Foundation B. local communities C. Voluntary clubs

15. Which service is NOT offered to cyclists?

A. picking up the bikes B. accepting responsibility for lost properties C. returning the cycle within two months

DISCUSSION

The modified version of the achievement test can be considered as practical, due to the clear instruction given with administration examples. As all Questions from 1 to 15 were provided with clear directions and they were easy to score (15 questions for 25 points, each right answer is applied for 1.6 points), the administrative issues did not emerge. Therefore, this test coincides with the description of Brown (2003, p 26) which means the practicality of the given test is achieved.

The required tasks were applied from both selected response and limited production questions: "gap-filling" and "matching" tasks were provided as well, which means the types of tasks were improved. As a result, the test validity could see improvement, because as Gronlund (1998) mentioned, the questions were meaningful and appropriate to the level of the student. Sentence completion tasks were given for the first four questions (Questions 1-4) in which the students were asked to fill the unfinished statements by jotting the words from the text. For example: Question 2 Sushi began to be served in _____ of educational settings.

This task assists to improve selective reading performance. According to Khalifa and Weir's (2009) classification, selective reading improves to grasp the meaning of the topics faster. Moreover, in this task, the number of words for answers was shown as well. This undoubtedly shows that the test reliability also was taken into consideration.

As for the next three questions, (Questions 5-7) the student was given TRUE/FALSE/NOT GIVEN task with providing the example: Question 6 "Raw shrimp is the choicest sushi so far."

The task distraction has covered the synonyms of the words, (a choicest and popular choice) which means the authenticity of the test is high enough. The learner can come across different types of synonyms in everyday reading. As Brown (2003) highlighted, the test offered various items which reflected the real-world. Furthermore, in the given sample, the words "eliminate" and "stop" were demonstrated as paraphrased words, in order to show that the information was TRUE. According

to the "Guidelines for preparing tasks of receptive skills" by Genesee and Upshur (1996) and Hughes (2003), "the questions follow the order of information". The data in the sample (in 1824 in Yohei Hanaya) was a hint about the location and order of searched information in the passage.

Questions 8 to 12 required learners to match the headings for the passages and skim the passages and identify the gist. The design of this task from Reading Passage II, as an example of an "even more sophisticated level" (Alderson, 2000, p. 218), was required for matching tasks.

Questions from 13 to 15 were multiple choice type questions and as Hugas (1989) mentioned: "the candidate provides evidence of successful reading by making a mark against one out of the number of alternatives" Moreover, multiple-choice questions provide high reliability, as they are easy to score. For example,

Question 14

Which organization usually prepares the drinks?

A. The British Heart Foundation B. local communities C. Voluntary clubs

It would be a good chance to check the learner's selective reading performance.

To influence positively to the learner's preparation, washback was given. As Brown (2003) explained the role of washback, it would motivate a student in further development in studies. The term washback did not exist in the previous achievement test (which was held on January 28, 2019, as a summative test). The student's result was returned simply through the grade in percentage, rather than lack of attempt to enhance the learner's comprehension. Giving washback to the learner enhances the learner's sense of being motivated, self-confident, having language ego, and strategic investment (Brown, 2003, p 38), it also helps to build a strong bond with the teacher. The learner begins to feel that there is always a chance to recognize and correct the errors.

I administrated the modified version of the test with my learner on February 18, 2019 just after four weeks. As the table above shows, the overall result was appropriate to the requirements of our curriculum, above 80% and from Reading they got 22%, compared to the previous score (January, 28, 2019), the percentage of result improved by 4%.

I told my colleagues about my learner's success and I had an attempt to explain that, the matter of my learner's higher result (22% out of 25% of Reading section) was achieved just by modifying the test. I showed every stage of the process, ranging from learner's profile to the last result. (February 18, 2019). The reasons for the achievement of the test, which will be shown below, were demonstrated to them via PowerPoint slides.

Although it was pretty tough to persuade some of my colleagues to modify the given UZTEA tests, they were at least aware of the reason for the low scores of learners. The point is that it is not always the teaching methods. Sometimes assessments require some changes/modifications as well.

CONCLUSION

The reasons for the improvement tend are as follows:

a) Test practicality: the time for reading section was prolonged, from 25 minutes to 30 minutes, similarly the number of questions were decreased, from 25 to 15, resulting in fewer questions but more time.

b) Test reliability: clear instruction was given by providing examples for each Reading section, while comparing the latest one. There were no examples for any question in the summative assessment.

c) Test validity: achieved through measuring the theme which was taught (the topic about food and transportation). Moreover, useful information was provided through the Reading passage and the questions were meaningful

d) Test authenticity: the questions can be seen in real life and question items were contextualized.

e) Washback: after taking the test, the learner was given feedback about the mistakes and offered to work hard on multiple choice questions, because two of their errors were from that task. After constructive feedback with explanations, they felt positively influenced.

f) The questions were only "selected-response" type, however "limited production tests" were added into the modified version

g) The given time was also changed, from 25 minutes to half an hour

h) The number of the questions faced a significant change, reduced from 30 to 15

i) Most importantly: INTENT was achieved and the RESULT improved

REFERENCES:

1. Alderson, J. C. (2005). Diagnosing language proficiency: The interface between learning and Assessment. London Press. Bachman, L., & Plamer, A.S (1996). Language testing in practice. New York: Oxford University Press.
2. Abdullaeva, Barno & Pulatova, Laziza & Ulmasbaeva, Malika & Azimova, Gulnara & Kenjaeva, Mukhayyo. (2020). Designing Materials for Teaching Adult Learners. International Journal of Psychosocial Rehabilitation. 24. 794-799. 10.37200/IJPR/V24SP1/PR201219.
3. Asrorov, I.F. (2020) "THE IMPORTANCE OF INDIVIDUAL FEATURES IN THE CLASSROOM," Central Asian Problems of Modern Science and Education: Vol. 2020 : Iss. 1 , Article 13
4. АРИПОВА, К. ЛИНГВОМАДАНИЙ КОМПЕТЕНЦИЯ ВА КОМПЕТЕНТЛИЛИК: ТЕРМИНЛАР-НИНГ МЕТОДИК ЛУТАТЛАРДАГИ ТАЪРИФИ. УЗМУ ХАБАРЛАРИ ВЕСТНИК НУУЗ
5. Khasanova, D. (2019). It is time to change the subject matter of English at vocational college EPRA International Journal of Multidiscipline Research
6. Monthly Peer Reviewed & Indexed International Online Journal India, 5(12) 48-51
7. Khasanova, D. (2019) Халқаро PIRLS баҳолаш дастурини ривожлантириш асосида ўқиш ва ёзув саводхонлигини ошириш Сўз санъати халқаро журнали 5 (10) 49-54 Bachman, L., & Plamer, A.S (1996). Language testing in practice. New York: Oxford University Press.
8. Brown, H. D (2000). Principles of language learning and teaching (4th ed.) White Plains, NY: Pearson Education.
9. Brown, J. D. (1998). New ways of classroom assessment. Alexandria, VA: Teachers of English to Speakers of other languages.
10. Chun, C. (2005). An analysis of a language test for employment: The authenticity of the Phone Pass test. Language Assessment Quarterly, 3, 295-308.
11. Eshboyeva D. A. Effective ways in teaching English language //Актуальные проблемы гуманитарных и естественных наук. 2014. №. 8-1. С. 312- 313.
12. Eshboeva D. A. THE ESSENTIAL FEATURES OF THE LANGUAGE\ TEACHING METHODOLOGY // Modern Science. 2016. №. 6. С 77-79.
13. Khasanova, D. (2020). The features of conducting Business English as an ESL ERPA International Journal of Research & Development International Online Journal 5(3) 674-676
14. Khasanova, D. (2020). Conducting the lessons in three different teaching conditions «Роль иностранных языков и информационно-коммуникационных технологий в подготовке молодых учёных» 236-331
15. Khasanova, D (2019). Cultural stereotypes of Asian students in TESOL Prospects of world Materials of the XV international scientific and practical conference 10, 54-59
16. Khasanova, D. (2019) Does the way of assessing affect the result of testing? "Trends in Teaching Foreign Language: achievements and Solutions" International Conference 84-86
17. Khasanova, D. (2019). The impact of effective teaching Grammar in CLT Nauka I Praktiki 13-14
18. Khasanova, D. (2020). The features of conducting Business English as an ESL ERPA International Journal of Research & Development International Online Journal 5(3) 674-676
19. Hughes A. (1981). Conversational cloze as a measure of oral ability. English Language Teaching Journal 35:161-8.
20. Ikanova, L. (2019). THE IMPACT OF THE MATERIALS DEVELOPMENT TO THE IMPLEMENTATION OF THE LANGUAGE PLANNING AND POLICY. Electronic journal of actual problems of modern science, education and training, 7-13
21. Swain, M. (1990). The language of French immersion students: Implications for theory and practice. In J.E. Alatis (Ed.) Georgetown University round table on languages and linguistics (pp.401-412). Washington. DC: Georgetown University Press.

22. McNamara, Tim. (2000). *Language testing*. Oxford: Oxford University Press
23. Mousavi, Seyyed Abbas. (2009). *An encyclopedic dictionary of language testing* (4th ed.). Tehran: Rahnama Publications
24. Meaney, K. S., Bohler, H. R., Kopf, K., Hernandez, L., & Scott, L. S. (2008). Service-learning and pre-service educators' cultural competence for teaching: An exploratory study. *Journal of Experiential Education*, 31189–208. doi:10.5193/JEE.31.2.189
25. Mo'minova, D. A. (2019). Translation problems of some English expressions into uzbek in intercultural communication Buxoro Davlat Universiteti ILMIY AXBOROTI, 284-288
26. Molee, L. M., Henry, M. E., Sessa, V. I., & McKinney-Prupis, E. R. (2010). Assessing learning in service-learning through critical reflection. *Journal of Experiential Education*, 33, 239–257. doi:10.5193/JEE33.3.239
27. Paoletti, J. B., Segal, E., & Totino, C. (2007). Acts of diversity: Assessing the impact of service-learning. *New directions for Teaching and Learning*, 111, 47–54. doi:10.1002/tl.285
28. Ramsdell, L. (2012). Reciprocity: The heart of service learning. In J. H. Westover (Ed.), *Academic service-learning across disciplines: Models, outcomes and assessment* (pp. 9–16). Champaign, IL:Common Ground.
29. Rymes, B. (2002). Language in development in the United States: Supervising adult ESOL preservice teachers in an immigrant community. *TESOL Quarterly*, 36, 431–452. doi:10.2307/3588420
30. Simons, L., & Cleary, B. (2006). The influence of service learning on students' personal and social development. *College Teaching*, 54, 307–319. doi:10.3200 /CTCH.54.4.307-319
31. Stukas, A. A., Snyder, M., & Clary, E. G. (1999). The effects of "mandatory volunteerism" on intentions to volunteer. *Psychological Science*, 10(1), 59– 64. doi:10.1111/1467-9280.00107
32. TESOL. (2010). TESOL/NCATE standards for the recognition of initial TESOL programs in ESL teacher education. Retrieved from www.tesol.org/advance-the-field/standards/tesol-ncatestandards-for-p-12-teacher-education-programs
33. Resolution № 272 of Cabinet of Ministry of Republic of Uzbekistan on 1996, July 31 "On the Establishment of Banking College".
34. Ulmasbaeva, M. A. (2020). THE EFFECTIVENESS OF USING TECHNOLOGIES IN TEACHING AND LEARNING FOREIGN LANGUAGES. In ПРИОРИТЕТНЫЕ НАПРАВЛЕНИЯ РАЗВИТИЯ НАУКИ И ОБРАЗОВАНИЯ (pp. 176-177
35. Джумаева Н. Э., Сохибов А. Р. Педагогические термины и понятия //КАРШИ Каршинский государственный университет. – 2014. – Т. 70.
36. Шамсегова Б. Р. МЕТОД ПРОЕКТОВ НА УРОКАХ АНГЛИЙСКОГО ЯЗЫКА //Актуальные проблемы гуманитарных и естественных наук.– 2019. – №. 2. – С. 96-99.