Dilbar KHASANOVA
INTERNATIONAL ISLAMIC ACADEMY OF UZBEKISTAN, EFL teacher of the department classical oriental literature and source study.
khasanovadilbar88@gmail.com, A.Kadiri, Tashkent, 100011, Uzbekistan.

THE REVISION AND MODIFICATION OF TESTS FOR LYCEUM STUDENTS IN B1 LEVEL.

The following article investigates the achievement test, which was designed by Uzbekistan Teachers of English Association (UZTEA) administrations for Year 2 students of lyceums. The target of this research is to discuss the test constructions and suggest more reliable version of them if the reason of unsatisfactory results depends on principles of assessment. The importance of the Test Modification Project for English language teachers is that the result of tests tends to be under the impact of test quality. In order to make it clear, the differences between ready given tests and modified versions will be discussed according to test specifications through following steps

Key words: *Principle, testing, modification, achievement tests, skills, test usefulness, validity, practicality, authenticity of test*

Introduction

Working as a teacher for eight years tends to give me an opportunity to make a sense about what is teaching. This conception does not simply mean to conduct the lesson by informing the learners, but also it demands a tremendous analysis what you are 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 is here, how the assessments are designed and how valid and reliable they are. Sometimes the results could be unsatisfactory due to the assessment reliability, even though valuable lessons are held. Therefore, I made a decision to analyze the test results of one of my learners, whose name is Gafurjnova Charos. A level of 17-year-old student of Year 2 at lyceum is A2, according to CEFR.

Her background knowledge is based on school materials and currently General English is being taught. She was accepted according to: state-mandate assessment (15 Grammar based tests) and her last result: 79% out of 100 % (January, 2019)

The reason of choosing this student among groups was the unchangeable proportion of results, 78% and 77% in the 1st week of December, 2018 and in 4th week of January, 2019 in turn. (Table 2). Although her motivation, attendance and active participation remained consistently within two months.

A description of the educational setting

As for educational setting, where this learner is studying, which was founded on the basis of 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 field of banking, finance, accounting and insurance. The students are accepted according to the state mandatory exams on Mathematics, English and native language, by 15 questions for each subject. The correct answer is signed by 2 points and overall score must be 90, while passing score is changeable. According to performance of pupils it may fluctuate. In 2017, the highest score was 68point out of 90 and 322 candidates were accepted and they have been mastered by 57 qualified teachers. The number of graduates, who applied 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 is commenced to teach from B1 level according to CEFR requirements and maintains being taught during 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.

Methods of most recent achievement test

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 every 7th week of the semester. As "the appeal of formative assessments is growing" (Ross, 2005), it is urgent to analyze the achievement tests and modify them, if it is necessary. The satisfactory results of formative tests (higher than 70,1% out of 100%) could lead to coming up 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 to learner’s results, there was a marginal change (78% and 79%) in performance:

Table 2

Test date	Type	Listening (25)	Reading (25)	Writing (25)	Speaking (25)	Overall (100)
December, 3.2018	Achievement, Formative	22	20	16	20	78% “4”
January 28,2019	Final, Summative	22	18	17	22	79% “4”

When it comes to implications of **Reading tests**, the lower percentage of result can affect instructions of classroom, as the teacher can mainly focus on the weak points and time management can

suffer, because of given too much time to the mistakes of the learner. Nevertheless, these progress checking tests, particularly Reading part, 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 of usefulness and low demand to maintain these Reading section tests and some test specifications will be discussed in following paragraph.

Critique of an existing test

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 unreachable goal to progress, due to quantity of lessons (only once per week). In order to give clear picture of achievement test, I would like to focus on the description of recent one, conducted on January 28, 2019. There are 25 questions for each listening and reading comprehension, which were followed by a 20-minute-writing task and the cue-card question to measure speaking skill. However only 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. (See the Appendix 2)

Table 3

Skills	Types of the tasks			Assessment	
Reading	Multiple choice question	True/ False/ Not Given	Overall number of tasks	Overall point (100%)	Given time (80minutes)
Given	16	9	25	25 %	25 min
Found	11	7	25	18	25

The connection between test and principles

The reading part of the assessment was much more inadequate for test principles, than other sections. Therefore, only Reading Passages are aimed to discuss. 10 question items out of 25 from the sample of Reading Test consisting of selected response questions Multiple-Choice and other 15questions are True/ False Tasks. All of them will be analyzed according to principles of language assessment. Reading passage will be provided. (See the Appendix 2)

Starting descriptions with criterion of an effective test, the Reading Section tests above were not “made from assessment results are appropriate, meaningful in terms of purpose of assessment” (Gronlund,1998, p 226), which means ***validity*** of the given test is not high. Moreover, content validity of the Test is somewhat invalid, because the students learnt “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 improving, such tests are believed as valid one, as in both B1 certificate and State Mandatory exams the students will be familiar with these multiple choice and True/False tasks, such as: **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 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. Besides that, the test lacks clear instructions for Part 1 and Part 2 with absence of an example. In **Questions 9, 12,15** and **16,** the lines are shown while there are no any 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) in order 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 narrative and analysis based topic, however 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 high likelihood to being an authentic test. As the question items of Reading are contextualized and interesting topics were chosen as a component of test and source was taken from real-world materials, the test can be authentic. **Question 23** can be example for 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.

Strengths and weaknesses of the test

Despite these bias, there are some strengths(S) of this assessment as well and they will be given with contrasting weaknesses(W).

1. S) As the test requires high Reading comprehension, it can suit both CEFR requirements and curriculum standards, which demand the learner to reach B1 level 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 certificate of B1 level, which is conducted by State Testing Center.

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

2. S) If a student is not capable to pass 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, as a traditional assessment. Therefore, focusing on discrete options will be a thorough preparation for state- mandatory exams. It would not be 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 in appropriate level and, which is an issue at this Reading test. If the unfamiliar Vocabulary and Tenses used in the passages were mastered in previous lessens, it would have suited to the Curriculum requirements.

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

W) Apart from level of difficulty and inappropriate level, the test includes only selected response questions, any type from "Limited production Tasks, such as, Gap-filling, short answer, dialog completion" (H.D. Brown 2006) which shows one of the undeniable issue. Even though the question type is remained the same,

say, Multiple-choice, it needs to simplify. It can check either Grammar or Vocabulary through Form-focused criteria.

Modified version of chosen test

Suggestions for improvement

Unless all five test principles are not applied to the evaluation of any classroom tests, such kind of tests tend to be inappropriate. For this reason, my suggestions will hopefully be how to cover to all requirements.

As the given test's validity and reliability are low enough, my concern of suggestions will focus on improving 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 teacher is "not to be 'tricky', but to encourage them and bring out the best in their performance". In order to gain this point, firstly the test itself must be at appropriate level. Even though the Reading passage remains the same, it would maintain 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 the curriculum requirements. According to the curriculum, the reading text should coincide what it was already taught. Therefore, as Hugas (1989) mentioned, prior to selecting the text there are some guideline which should be obeyed, such as: not selecting repeatedly 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 too academic, make sure whether only reading comprehension is tested or not.

Concerning reliability of the discussed test, it also be analyzed due to the time. In order to provide 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.

A design of modified version of test with new instructions.

A proper way to commence giving modified version is shortening the numbers of passages, questions and alternatives. It will be two passages, 15 question stems with three (A, B, C) options instead of 25 items with five distractors. Part I, due to the topic of it, it will 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 also may remain as the topic to check comprehension. Both the reading passages will be given to apply, after introducing 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 in order to impact the cost of 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 opposite direction?

Other cyclists B. Passengers C. Vehicles

Answer: C

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

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

*14.*Which organization usually prepare 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

Findings of modified version of the test

The modified version of the achievement test can be considered as *practical*, due to its given clear instruction with examples for administration. As all Questions from 1to 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 point), the administrative issues were not emerged. Therefore, this test coincides 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 the improvement, because as Gronlund (1998) mentioned, the questions were meaningful and appropriate to the level of the student. Sentence completion task were given for the first four questions, from Question 1 to Question 4, 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 the words for answers were shown as well, which shows that the test *reliability* also was taken into consideration.

As for 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, (choicest and popular choice) which means the *authenticity* of the test is high enough. The learner can come across different type of synonyms in everyday reading.as Brown (2003) highlighted, the test offered the items, which reflected 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 "the questions follow the order of information". The data in the sample *(in 1824 in Yohei Hanaya)* was the hint about location and order of searched information in the passage.

Regarding Questions from 8 to 12, to put the headings for the passages was required, where the learner had to skim the passages and identify the gist. The reason of designing this task from the Reading Passage II, "even more sophisticated level" (Alderson,2000, p. 218) was required for matching task.

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 number of alternatives" Moreover, multiple choice questions provide high reliability, as it is easy to score. For example: **Question 14**

Which organization usually prepare 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.

In order 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 previous achievement test (which was hold on January 28,2019 as a summative). 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 except giving an opportunity to be motivated, self-confident, having language ego and strategic investment (Brown, 2003, p 38), it also helps to build the strong bond with the teacher. The learner begins to feel that there is always chance to treat the errors.

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

REFERENCE:

1. Alderson, J. C. (2005). Diagnosing language proficiency: The interface between learning and Assessment. London Press.
2. Eshboyeva D. A. Effective ways in teaching English language //Актуальные проблемы гуманитарных и естественных наук. – 2014. – №. 8-1. – С. 312-313.
3. Шамсематова Б. Р. МЕТОД ПРОЕКТОВ НА УРОКАХ АНГЛИЙСКОГО ЯЗЫКА //Актуальные проблемы гуманитарных и естественных наук. – 2019. – №. 2. – С. 96-99.
4. 4. Джумаева Н. Э., Сохибов А. Р. Педагогические термины и понятия //КАРШИ: Каршинский государственный университет. – 2014. – Т. 70.
5. 5. Eshboeva D. A. THE ESSENTIAL FEATURES OF THE LANGUAGE TEACHING METHODOLOGY // Modern Science. – 2016. – №. 6. – С. 77-79.
6. Bachman, L., & Plamer , A.S (1996). Language testing in practice. New York: Oxford University Press
7. Brown, H. D (2000). Principles of language learning and teaching (4th ed.) White Plains, NY: Pearson Education.
8. Brown, J. D. (1998). New ways of classroom assessment. Alexandra, VA: Teachers of Englis To Speakers of other languages.
9. Chun , C. (2005). An analysis of a language test for employment: The authenticity of the Phone Pass test. Language Assessment Quarterly, 3, 295-308
10. Hughes A. (1981). Conversational cloze as a measure of oral ability. English Language Teaching Journal35:161-8
11. 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.
12. Resolution № 272 of Cabinet of Ministry of Republic of Uzbekistan on 1996, July 31 "On the Establishment of Banking College"
13. Resolution №143 of Cabinet of Ministry № "On measures to accelerate equipping classes Foreign languages of educational institutions in the Republic with the modern information- communication equipment, technical means of teaching in 2013-2016". May 23,2013