

TEST MODIFICATION PROJECT

Dinora Sodiqova,

Senior Teacher of Nordic International University,

Tashkent, Uzbekistan.

Tel: +998974126779

E-mail: dinorasodikova1990@gmail.com

d.sodiqova@nordicuniversity.org

Annotation. Traditionally teaching processes are followed by testing the acquired knowledge through courses. Designing test materials with appropriate questions, distractor along with precise keys are considered to be responsible duty which requires meticulous analysis. However, experts might face with issues such as two correct keys or inappropriate context or content of the questions. Addressing these testing problems the following article investigates the concerns and the modification of one particular test that has been used in one of the schools of Uzbekistan.

Key words: a distractor, modification, test materials, MSQs (Multiple Choice Questions), stems.

Introduction. The following Test Modification Project investigates the test provided by a school teacher as a final midterm assignment which is going to be analyzed and modified according to the theories and five principals (Brown, 2010) of language testing an assessment.

Three main parts are included in the work:

- ✓ Learner assessment profile
- ✓ Critique of the existing language assessment
- ✓ The modified version of the selected assessment

The chosen student at her will (Appendix 1), “Mary” is a 15-year-old girl, with high ambition of studying. Being positively enthusiastic to education and not indifferent her achievement along with errors which occurring in the learning process, it was she who

suggested to participate in the project when all the students of the 10th grade were proposed to take part in (Appendix 1). Her first language is Uzbek, however, being Russian her second language she studies in a class where Russian is the language of education. She has been studying English thoroughly for 2 years at different language learning centers, besides, since 5th grade that language has been taught as a foreign language compulsory at school #163 in Chilanzar district of Tashkent City where as other secondary schools English is taught three times a week and each class lasts 45 minutes. Additionally, the English language acquaintance of Mary's family members is considered to be one more supportive point in terms of acquiring the language. Especially, "they are so much helpful as they speak common daily life conversation in English with me" states Mary, that indicates her speaking skills are enough to make daily conversations in this language.

Regarding to her grades from the progress checks during the first and second term after finishing each Unit of their Textbook at school, it can be inferred that she is A student with almost high scores. However, occasionally her scores are below expectations which are making her concern a little. Actually, it is not the case for her to be so worried which I am going to reveal the reasons of the agenda in the section of "critiques of the existing assessment" later.

Taking her grade (10th) into consideration Mary is to be B1 level according to the CEFR. In comparison with her classmates Mary's knowledge of English seems to be sufficient to take the tests which her English teacher provided.

In the syllabus of a secondary public schools there are 7 Progress checks (they were identified as Control work 1,2, 3, 4,5,6,7) as a form of checking and revising the themes have been taught, 2 Progress Checks for each term with exception for the second term. As Term 2 is considered to be the shortest term, only one progress check in the middle of the term was planned to be held. Moreover, 4 Final tests are included at the end of each term for the one academic year. According to the curriculum (Appendix 2) designed by the Center of Ministry of

Education four, both receptive and productive skills are suggested to be improved which is highly beneficial to be undertaken in the assessment process as well.

The Progress Check Test 4 (Appendix 2) designed to check the achievements of Unit 6, including eight lessons, is planned to be taken in the second week of the second term. The Progress Check test which was suggested to be analyzed by Mary was developed by the administrator of the blog Uzteachers named Hasanboy Rasulov who is a senior English teacher as well. The progress check test (Brown, 2009, p. 9) intends to check a student's both receptive and productive skills. Although the participant's results were high as she expected (she got 5-), Mary has confusion in terms of reliability and practicality of the questions as she stated that it was challenging to realize what she is expected to do (Appendix 3). The explanation which Mary was provided by her school English teacher did not satisfy her as she was delighted and eager to participate in the project.

Definitely, the syllabus alongside with assessment planning ought to coincide with the curriculum which was designed by Ministry of Education the participant did not have a chance to take Placement test at school in the beginning of a school year. However, it would be practical to implement a placement test before starting every academic year and accordingly to divide the classes into groups. As it can be inferred from Mary's claims the case is different in her school, precisely in her class.

The Progress Check (it is "Control work 4" is in the syllabus) is designed by Hasanboy Rasulov and was practiced in a school teacher's lesson with the permission of the author which consisted of items on the themes taught for two weeks in the first half of the second term. It includes overall 3 parts with 20 items, that is 10 items for Vocabulary and Grammar section, as a form of Multiple Choice Questions (MCQs). Each item possesses four different distractors from which must be chosen one correct answer. The second part is listening with 7 stems requiring to identify if they are True or False. The final part of the test is based on a reading material along with 3 items, 2 of them contains 5 distractors while one has only four distractors.

The required time to fulfill the test is set as 45 minutes. A teacher is provided with keys to check and assess them, however, neither rubric nor exact points for the items were demonstrated for a test checker which may affect the reliability of the test.

For this particular test the most important factor is considered to check the unit which students have finished studying recently. Furthermore, the appropriateness was taken into consideration as I had an opportunity to check whether it coincides with the level of the students of 10th grade or not. Due to the fact that the second part, namely listening part can support all principals of test designing.

For this purpose, I used a website www.textinspector.com where not only the level of the test but also the other qualities such as readability or syllable count might also be analyzed. According to CEFR in terms of receptive skills (i.e. reading and listening) B1 candidates should meet the following requirements:

B1	Can understand straightforward factual information about common everyday or job related topics, identifying both general messages and specific details, provided speech is clearly articulated in a generally familiar accent.
	Can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure etc., including short narratives.

Reading descriptor:

B1	Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension.
-----------	--

It can be evident that the test itself dependable regarding to its reliability (Bachman,1990, Brown, 2005, Fulcher & Davidson, 2007 cited in Brown, 2010, p. 27). Besides, the test objective which have predetermined responses and subsequently does not require individual judgement as one more reliable indicator. Moreover, not only the test items are appropriate with the fixed time to be solved and selected but also it can be scored easily as Mousavi (2009) stated in the criteria of being a practical one. Undoubtedly, the clearer items and keys the fewer test- checkers are in need of the process, either one examiner or test-taker is sufficient to assess and evaluate the results of the test- taker.

Except the criteria which have been mentioned, validity of the test can be analyzed. Since the test is planned to be taken in 45 minutes, the number of items strictly controlled in order to end it on time. Yet, Weir (2005), Broadfoot (2005), and McNamara questioned the test validity emphasizing that is “the matter of degree”.

However, there are number of drawbacks of the test can be observed, such as the test, apparently, does not cover all productive and receptive skills. That is, the test constitution is limited with grammar and reading MCQs along with listening matching assignment. The other productive skills together with speaking skills were neglected while creating the test. Oller (1979) argued that language competence is a unified set of interjecting abilities that cannot be tested separately (Oller, cited in Brown 2010).

Second issue is closely connected with authenticity of the discourse as the text and stems are not related to the real life materials but rather taken from the textbooks specially prepared for language learning purposes.

Although it was mentioned that due to the objectivity test may earn the reputation of reliable, test distractors and instruction are not carefully organized as it was suggested by Brown (2009) in the chapter which can occur measurement errors. In the sample 1, the task of the Part 1 seems to be not appropriately was constructed.

Control work 4

1. Choose one correct answer

1. Students _____ behave well when they're in class.
A) must B) shouldn't C) need D) have to

(Sample 1)

It can be inferred from the sample only one of the distractors is demanded to be a key response, yet two of the distractors “must” and “have to” might be possible keys (Mann & Taylore – Knowles, 2008).

7. They _____ come to the cinema with us.
A) doesn't B) not C) didn't D) do

(Sample 2)

The similar case is observable in the seventh stem as long as two keys are suitable. Once it is scrutinized attentively it becomes obvious that both distractors C and D are likely to be potential keys. An extensive body of literature exist on the topic, for instance, according to Mann and Taylore- Knowles (2008, p.8) and many other grammarians in a statement the auxiliary “do” is used for emphasis, meanwhile, “did not” may be possible as a role of negative form in the past simple. Concerning on the content the test questions it is validity is being questioned rudely, distractors and keys are to be to studied and analyzed thoroughly as Hughes (2003) mentioned.

Modifying process. As listening section of the test meets all requirements of test designing, in this part grammar and reading MCQs are going to be optimized in accordance with justifications which are provided via reliable sources.

The following modifications are taking place in:

1. Replacing 6 of the stems with Vocabulary aligned with appropriate options.
2. Replacing the distractors adding context in the Grammar part.
3. Providing instruction for the reading part.
4. Designing equal number of distractors for the discourse taking the level of the students into consideration.

As Brown (2010) states in the cases when unit is being tested to identify how well students achieved the goals and objectives content-related validity is observable and also Hughes (2003, p.173) mentions “tests where teaching objectives or the syllabus list the grammatical structures to be taught, specification of content should be quite straightforward”. However, unless adding vocabulary features having been studied both Grammar and Vocabulary in the unit and the test is covering only Grammar, it can be claimed that content- related validity

suffers in the context. Consequently, a learner's achievement may not be completely analyzed. Therefore, I have chosen to leave 4 items (items number 1,2,3,4) based on grammar but with corrected distractors and extending context, that is now 4 distractors from which a candidate is able to choose only one as it was instructed in the original version of the test without any other correct options to improve practicality of that.

Item #1. Students _____ behave well when they're in class. (original)

A) must B) shouldn't C) need D) have to

Certainly, context can be added in the stem to avoid the other possibilities as Hughes (2003) suggested. Yet, if to follow the instructions still we cannot select one of the distractors as a key, as there are two possible responses in the line: a) must d) have to. In order to avoid two responses, I decided to change the distractors as well.

Item #1.. (modified)

A) Need B) ought C) have D) must

In this option the absence of the preposition 'to' after 'ought', 'have' makes the key 'must' unique. Besides, the context is indicating to the obligation which must be done. Similar shifting has been carried out with the other items of the grammar related question up to item 4.

The rest of the Progress Check 4 (it would be better to rename it as Progress Check4 instead of Control work 4) supplemented by Vocabulary items from 4 till 10.

Item # 5. Which of the following answers can be synonym for the phrase "to kill time"?

Distractors: A) Shoot time B) pass the time C) waste time D) be out of time

Regarding the level of the students, stem can slightly be challenged to reach face validity as it was quoted from Gronlund (1998) the criteria in terms of assessment are fairness, relevance and benefits of improving studying. Rather than simply requesting "What does 'kill the time' mean?", it is effective to implement specifically constructed stems. Once the distractors are scrutinized, it becomes apparent that other than "pass the time" fails to be key. This type of MCQ is identified as recognition of synonyms (Hughes,2003). Both the phrase

asked in the stem and the distractors are familiar to the learner from the unit which they have been taught according their textbook in Unit V. Time management (Rashidova, 2018) during those 2 weeks which may indicate content- validity of the items. Regarding the content item number 9 is also can be classified as item number 5 (See the Table)

Discussing another question, item # 6 reveals the other features of the first part of modified test:

Item # 6. Hooray, dad, we're happy to see you. Have a seat. You are _____ for lunch.

Distractors: A) out of time B) lack of time C) just in time D) waste time

Recognizing appropriate word for context (Hughes, 2003, p.183) is also believed to be one of the efficient way of revising the material, additionally coming across with the word in the context may provide the authenticity of the task. The statement given in the stem is related to a daily speech of the speaker which confirms the real –life connection of the item. The length of all the distractors are equal, connected to the theme which is being questioned which provides reliability of the item. The unlike options for the stem does not distract a candidate from the knowledge that she or he gained in the process of learning. Correspondingly, items number 8 and 10 are also designed in the same way with item number six.

Subsequently, one more differently constructed item exists in the test which should not be prevented to look over, that is item number seven in which candidate is challenged not only in terms of brilliance of his or her memory but also the rate of critical thinking skills is intended to be revealed.

Item # 7. How do you explain the word " priority"?

Distractors: A) Something that is very important and must be done first

A) a log or diary of a sequence and duration of activities engaged in by an individual

B) a genealogical diagram of a people who live together

C) a person who starts first activities

According to Hughes (2003) this type of testing “recognizing definition of Vocabulary “does not always practical because of the reason that not all words can possess clear and specific meaning, yet, that type of tests works well with “a range of lexical items” (p.183). The stem as well as the distractors were introduced during the classes, apparently the item meets the requirements as it might be practical as does not create difficulties with assessing (should be pointed that all definitions were taught according to their textbook). Aiming at to compare two option of the tests, existed and modified, A Table below created.

Item #	Original stems and distractors	Modified stems and distractors	Changes/ Justification
1.	<p>Stem: Students _____ behave well when they’re in class.</p> <p>Distractors: A) must B) shouldn’t C) need D) have to</p> <p>Key: A) must</p>	<p>Stem: Students _____ behave well when they’re in class. If they don’t their parents will be fined</p> <p>Distractors: A) need B) ought C) have D) must</p> <p>Key: D) must</p>	<p>distractors have been changed, context added</p>
2.	<p>Stem: We _____ organize our portfolio. At the end of the term it will be marked. (original)</p> <p>Distractors: A) must B) need C) have to D) may</p> <p>Key: A) must</p>	<p>Stem: We _____ organize our portfolio. At the end of the term it will be marked.</p> <p>Distractors: A) has to B) have to C) could D) need</p> <p>Key: B) have to</p>	<p>distractors have been changed to avoid confusion</p>

3	<p>Stem: There are plenty of tomatoes in the fridge. You ____ buy any.</p> <p>Distractors:</p> <p>A) mustn't B) needn't C) shouldn't D) may not</p> <p>Key: B) needn't</p>	<p>Stem: There are plenty of tomatoes in the fridge. You buy any.</p> <p>Distractors:</p> <p>A) Must B) can C) needn't D) can't</p> <p>Key: C) needn't</p>	<p>distractors have been changed to avoid confusion</p>
4.	<p>Stem: Drivers . . . stop when the traffic lights are red.</p> <p>Distractors: A) must B) need C) mustn't D) have to</p> <p>Key: A) must</p>	<p>Stem: Drivers . . . stop when the traffic lights are red. Otherwise, bad accidents may happen.</p> <p>Distractors:</p> <p>A) must B) mustn't C) need D) needn't</p> <p>Key: A) must</p>	<p>Order of the item, distractors have been changed, context added</p>
Items for checking materials on Vocabulary			
Item #	Original stems on Grammar	Modified stems and distractors	Justification
5	<p>Stem: If you want to learn to speak English fluently, you . . . to work hard.</p> <p>Distractors: A) must B) need C) should D) may</p> <p>Key: B) need</p>	<p>Stem: Which of the following answers can be synonym for the phrase "to kill time"?</p> <p>Distractors: A) Shoot time B) pass the time C) waste time D) be out of time</p> <p>Key: B) pass the time</p>	<p>Content – validity suffers if just grammar, vocabulary materials also have been studied</p>

6	<p>Stem: Carla _____ to the radio all morning.</p> <p>Distractors: A) listening B) heard C) listened D) is listening</p> <p>Key: C) listened</p>	<p>Stem: Hooray, dad, we're happy to see you. Have a seat. You are _____ for lunch.</p> <p>Distractors: A) out of time B) lack of time C) just in time D) waste time</p> <p>Key: C) just in time</p>	<p>Content – validity suffers if just grammar, vocabulary materials also have been studied</p>
7	<p>Stem: They _____ come to the cinema with us.</p> <p>Distractors: A) doesn't B) not C) didn't D) do</p> <p>Key: C) didn't</p>	<p>Stem: How do you explain the word " priority"?"</p> <p>Distractors: A) Something that is very important and must be done first B) a log or diary of a sequence and duration of activities engaged in by an individual C) a genealogical diagram of a people who live together D) a person who starts first activities</p> <p>Key: A) Something that is very important and must be done first</p>	<p>Content – validity suffers if just grammar, vocabulary materials also have been studied</p>
8		<p>Stem: I _____ before</p>	<p>Content –</p>

	<p>Stem: I like this song. _____ do I.</p> <p>Distractors: A) Either B) So C) Neither D) Nor</p> <p>Key: B) So</p>	<p>the test, so I couldn't answer for the last 5 questions.</p> <p>Distractors: A) killed time B) wasted time C) ran out of time D) high time</p> <p>Key: C) ran out of time</p>	<p>validity suffers if just grammar, vocabulary materials also have been studied</p>
9	<p>Stem: We _____ them at eight o'clock.</p> <p>Distractors: A) meet B) 're meet C) 're meeting D) is meeting</p> <p>Key: C) 're meeting</p>	<p>Stem: Which of the phrases below can be similar to the "have a rough time"</p> <p>Distractors: A) have wastage of time B) have a short of time C) run out of time D) have a difficult time</p> <p>Key: D) have a difficult time</p>	<p>Content – validity suffers if just grammar, vocabulary materials also have been studied</p>
10	<p>Stem: He _____ know how to spell it.</p> <p>Distractors: A) doesn't B) hasn't C) don't D) does</p> <p>Key: A) doesn't</p>	<p>Stem: It's important to come _____ for a job interview. It shows a candidate's responsibility.</p> <p>Distractors: A) about time B) on time C) in short of time D) hard time</p> <p>Key: B) on time</p>	<p>Content – validity suffers if just grammar, vocabulary materials also have been studied</p>

Table 1. Comparison of the existing and Modified Test.

Eventually, regarding the needs of the participant who complained on the failure of the comprehension regarding the third part of the Progress Check 4, I opted that it would be reasonable to provide a clear and specific instruction for the discourse being asked.

The source of the text has not been referenced by the test- maker, but it is not the issue which is going to be investigated in the project. The agenda is the absence of the instruction which concerned my participant. Bachman (1990) and a number of experts claimed that the appearance whether it is stem, distractor or instruction can definitely affect the test- takers result which consequently make “student- related unreliability” (Brown, 2010).

Instruction for the reading part:

Part 3. Reading. Read the given text carefully and choose one of the options for the questions 18,19, 20. Only one answer is possible.

In an interview yesterday Mr. Wilson was questioned about the harmful effects of horror movies on teenagers. He argued that such effects were often exaggerated and claimed that other types of films were far more dangerous for young people. When asked to prove this, he pointed out that horror films were often set in unreal situations and were clearly not to be taken seriously. In contrast, he claimed that films showing violent crime were often set in everyday life, and were therefore more damaging.

18. In Mr. Wilson’s opinion, horror films _____.

- A) cost more than other kinds of films.
- B) are more popular among the elderly than among the young.
- C) should be banned altogether.
- D) are less damaging to young people than films of violent crime.
- E) have recently ceased to appeal to the young.

19. For Mr. Wilson the main difference between a horror film and one showing violent crime is that the former _____.

- A) is mainly concerned with everyday situations.
- B) is liked by the young, and the latter by the old.
- C) is unrelated to real life, whereas the latter is.
- D) is less expensive to produce than the latter.
- E) rarely receives any attention from the young.

20. The interviewer wanted to find out whether _____.

- A) young people were being harmed by horror films.
- B) Mr. Wilson preferred horror films to films of violence.
- C) people were seriously objecting to horror films.
- D) the effects of crime films were being exaggerated.

www.hasanboy.uz saytidan olindi

Implementing process. Intending to find out the efficiency and whether modified test meets the requirements of five principles which were suggested by Brown (2010) or not the modified option of the Progress Check 4, precisely the first and the third part of the test was implemented. For this purpose, I had to use www.google.com/forms because of the reason Mary was to go the region with family matters, eventually we needed to use internet tools occurred to be profoundly beneficial for us (Appendix 4). At first, I sent her the First Vocabulary and Grammar part in which she made no one mistake, following that she had reading part of the test later and that one also can be estimated as a successful attempt (Appendix 5). To assess the results and give proper mark a rubric (See Table 2) was developed. Although the test is summative assessment and results are assessed by the analytic rubric it would be efficient to integrate formative assessment with that as a form of constructive feedback if a mistake occurs in student's replies. In my participant's case there was no need to do that.

Rubric	Grammar and Vocabulary	Listening	Reading
A	9-10 Correct responses	6-7 Correct responses	3 Correct responses
B	6-8 Correct responses	5-4 Correct responses	2 Correct responses
C	5-4 Correct responses	3-4 Correct responses	1 Correct response

Table 2. Rubric for the Test

Having analyzed Mary's progress, I desired to find out her comments and opinions on the modified test. Consequently, a feedback survey was developed in the same website (www.google.com/forms) to scrutinize if the tests may be piloted in my own classroom with 10th graders in the future (Appendix 6). After reading her replies for the feedback which consists of five questions in a target language understandable for Marry, it can surely be confirmed the test is effective and does not question need of a learner.

Conclusion. Having thoroughly investigated the existing test addressing to the linguist's chapters and articles it can be concluded that designing a test is not a just killing the teacher's classroom teaching time. In order to achieve rocketed efficiency of the assessment several factors, such as practicality (Mousavi,2009), validity, reliability (Bachman,1990, Brown, 2005, Fulcher & Davidson) authenticity, impact and washback (Brown, 2010) need to be taken into consideration.

Analyzing the needs of the learner as it was suggested by Graves (2000), it was obvious that the existed test lack several qualities which were described five principals suggested by Brown. Keeping in my all those necessities I tried to modify the test looking over the features explained in Hughes's (2003) chapter.

After piloting the test results were examined with the rubric which was developed and constructive feedback provided in a positive way as to show the approval of her achievement (William, 2011). In turn, as a feedback Mary's answers were proof to the practicality, authenticity and validity of the test as primary consideration.

References:

Bachman, Lyle F.(1990). *Fundamental considerations in language*. New Oxford University Press

- Brown, H. D. & Lee, H. (2015). *Teaching by principles an interactive approach to language/ pedagogy*. New York: Pearson Education, 10 Bank Street.4 (pp.315, 426)
- Graves, K. (2000). *Designing language courses: a guide for teachers*. P.100. Boston; Heinle and Heinle Thompson Learning.
- Hughes, A. (2003). *Testing for Language Teachers (2nd edition)*. Cambridge: CUP. <https://dx.doi.org/10.1017/CBO9780511732980.012>
- Mann, M. & Taylore- Knowles, S. (2008). *Destination B2: Grammar and Vocabulary*. (2nd ed.) Macmillan Education.
- Oller, John W. (1983) *Issues language Testing*. Rowley, MA: Newbury House
- Rashidova F., et al. (2018) “English book pupil’s book for 10th grade pupils of secondary schools. Tashkent: “Uzbekistan”
- Sodiqova, D. (2022). LINGUISTIC CHARACTERISTICS OF PHRASEOLOGISMS INVOLVING CLOTHES NAMES IN ENGLISH AND UZBEK. *Oriental renaissance: Innovative, educational, natural and social sciences*, 2(Special Issue 28-2), 206-212.
- Sodiqova, S., & Sodiqova, D. (2022). REKLAMANING LISONIY VA NOLISONIY XUSUSIYATLARI. *Journal of Integrated Education and Research*, 1(4), 717-721.
- William. (2011). What is assessment for learning? *Studies in Educational Evaluation*.37(2011) 3-14. Retrieved from <http://www.elsevier.com/stueduc>