

CASE STUDY: ANALYSIS OF DESIGNING READING TESTS

Kuchkarova Yana Davidovna
yakuchkarova@gmail.com

ABSTRACT

Nowadays, research on teachers' abilities to design test items to assess the reading skills of learners is a common topic in the EFL field. The purpose of this paper is to present insight into designing, modifying, and piloting the reading test in the higher education establishment. Also, the article analyses the reading tests used to conduct the summative assessment. The results can help outline future directions in assessment practices.

Keywords: Test, reading, items designing, assessment.

Learner Profile

The learner who volunteered to participate in this project is a second-year student who transferred from the Foreign Language Institute because of a family condition. Oydina (pseudonym), is a 19-year-old young adult, who is passionately learning English since she was 13 attending different courses at various centers in the region. She is an ambitious, enthusiastic, and optimistic girl whose dream is to study MA in the USA. She holds an IELTS 6 score with the help of which she entered the institute with a high score in English. Her language skills were scored according to IELTS score bands as Listening 7, Speaking 5.5, Reading 6, and Writing 6. Even though her Speaking was marked as 5.5, her speech can be found quite fluent now which might have improved since she took the exam. She is interested in various disciplines such as business, international relationships, and technology. Her marks in her first and second terms are 4-5 according to the grading system in the republic where the highest is 5 and the lowest is 0. The student gets 0 if he/she doesn't submit his/her final on the subject. Her favorite subject in Term 3 was Integrated Skills as they worked on different projects and had fun with their groupmates doing homework and her least favourite subject was Philosophy, the purpose of which she does not understand. The sphere she wants to improve in reading is her vocabulary related to different spheres, as it prevents her from understanding the passages in international language tests such as IELTS and TOEFL fully because of her lack of vocabulary. Her favorite task on the subject Reading-Writing was to prepare an encyclopedia chapter for kids for what they read articles from a chapter on Children's encyclopedia, written in English, summarized the articles into 1-2 sentence-encyclopedia entries and designed an encyclopedia chapter for kids.

The Educational Setting

The University is a public university that accommodates more than 20000 students at sixteen faculties. The faculty Oydina studies is considered one of the most appealing branches of the University. Oydina is a student of this faculty since the 2021-2022 academic year. According to the curriculum, the second-year students of the branch of Foreign Language and Literature have had the subjects such as Philosophy, Foreign Language (German or French), Country-study, General Pedagogy, Integrated language skills, Reading and Writing, Language aspects (grammar, vocabulary) in the third term. Reading and Writing are conducted for 30 hours which would be 15 pairs a term, i.e., one week Reading and the other week is Writing, which means students have Reading classes every other week. Based on the syllabus provided by the State University which is considered the main headquarters which design syllabi for future

English teachers, the objectives for the subject Reading is to develop reading strategies, and the themes set for them are the following:

Objectives:

- Extensive reading (literary reading)
- Intensive reading (reading for information)

Reading strategies to be developed

- reading for main idea
- reading for gist
- reading for specific information
- guessing words from the context
- attentive reading

-speedreading

-scanning/skimming

The themes to be covered:

Reading strategies. Intensive and extensive reading

Reading Newspaper and journal articles

Reading Encyclopedia articles

Reading Reports, analysis, summary

Reading Scientific articles

Reading News reports

Reading Short news reports for university newspaper

Information on the most recent placement test

The most recent placement test, the learner has taken was transfer exams on the subjects that the learners did not have the opportunity to study in their previous institutions. The aim of this exam is to evaluate if the learner acquired the knowledge on the subject on the independent study base well enough to continue the education at university. The exams that the Learner took were on Literature and CEFR (Common European Framework of Reference) which used multiple-choice questions to evaluate the learner. It was decided to evaluate the Learner on these subjects because these subjects were chosen as optional subjects for first-year students, whereas Oydina did not have these subjects at Foreign Language Institute while she was studying there in Year 1. The results would not mean so much in deciding to place the learner in the groups. The only factor that influenced the placement was the fact that she studied in the Russian group at Foreign Language Institute and when transferring to State University she was put in one of the Russian groups existing in the faculty for the academic year 2021-2022 to continue her education as a sophomore.

Critique of an existing test/assessment

The test to be evaluated for this paper is a final summative assessment for the subject “Reading and Writing” for Year 2 students. As the platform that is being used for the final tests, supports only multiple-choice test types, it was decided to prepare multiple-choice question tests for the summative assessment. Twenty-five small texts taken from the book “Test Master” designed by Ataly Oguz were modified and used for the reading summative assessment. The questions were changed somehow by the teacher in order to prevent the learners from just copying and pasting answers from the original book that is available online and quite famous at local educational centers that prepare applicants for the university entrance exams. The summative assessment is an achievement test by nature to evaluate to what extent the learners reached the objectives set in the curriculum (Brown 2010. P.9). The students were given 25 mini texts per one-item multiple choice question for one hour with two trials. As it was their first exam on

the platform while staying at their homes the instructor decided to have two trials just in case some technological problems arise while they are doing the test. According to Brown (2010), assessment can be evaluated by the following criteria: practicality, validity, reliability, authenticity, and washback.

Practicality considers “costs, the amount of time it takes to construct and to administer, ease of scoring, and ease of interpreting/reporting the results” as Mousavi (2009) cited in Brown (2010). From this point of view, the test is medium practical as it let the instructor evaluate more than 200 hundred students at a time. Platform did the scoring itself and student knew their marks right after the assessment finished. However, administering was not easy as the administration of the University was not generous with time, to let the instructor learn to navigate the platform and not every student has a computer or a phone with internet access to solve the test which might reduce the practicality level. The given time for learners was one hour for 25 different mini-texts with 25 multiple choice items, in my consideration, is not practical as most of the learners might have needed more time as they used their mobile devices which will not show the whole text at a time.

Hughes defined backwash as the effect of testing on teaching and learning. It can be negative as well as a positive effect (Hughes 2003). The backwash effect of the test can be quite low as it did not give the feeling of accomplishment and satisfaction to the learner as it was mentioned by Oydina. Moreover, she finds the test quite old-fashioned and the test items do not require her to skim or scan as other international tests such as TOEFL or IELTS do.

Authenticity check explores if “a task or language sample is “real-world” or not” (Brown 2010, p.36). Authenticity in reading can evaluate text authenticity and task authenticity. Texts used in the reading are modified, as when they were checked through the website lextutor it showed that the texts mostly consist of 1-1000 words. Moreover, the student interviewed mentioned that she does not find the text real that she reads in real life. In terms of task authenticity, the test utilized only multiple-choice items that required us to choose one out of four options something that we do not do when we do a reading.

Validity is referred to as the criteria that examine if the test does what it is supposed to do if it checks what it is supposed to check (Hughes, 2003; Brown, 2010). Evaluating the validity of this summative assessment one can notice that it does not include the Writing module of the subject, only Reading. Another concern might be in terms of the content validity of the test. The content of the test does not reflect the content of the course. Last but not least, the test items were taken because they were easily available and the instructor did not think much about what each item is aiming to measure. Hughes (2003) suggests that a very good test would measure what it is supposed to measure.

Analyzing the reliability of the test, it can be noted as a medium. Being administered online without support and supervision the test did not do its job, i.e., it did not give the instructor an idea of who is excellent-good or poor in reading. The students had the opportunity to cheat, some brilliant students could not show their best because of technology problems and anxiety they had which will lower the reliability of this assessment. However, scoring reliability can be found high as the computer did all the counting.

Three strengths of the Reading test administered as a final summative assessment can be its scoring reliability, short texts as it is easier to upload them to the platform rather than long texts, and its practicality in administering at the moment the test was taken.

Three weaknesses of this test can be authenticity, validity, and washback as it was explained in the paragraphs above.

A modified version of the chosen test/assessment

As it is presented in Standards for the Assessment of Reading and Writing (2010) assessment shapes teaching, learning, and policy. Depending on how it is treated it will improve or diminish curriculum and instructions. According to Hamamrad 2017 **test** is “the process of evaluating learners' ability, intelligence, as well as indicating their weaknesses and strengths in a learned material”. It is mentioned that Reading is probably the most frequently tested skill as it is assumed to be the easiest skill to check. However, it has its own difficulties (Hamamrad 2017).

Being a combination of two sub-skills “micro-skill” such as processing letters, words, orthographic patterns, recognizing word classes like nouns, verbs, etc. and understanding systems like tense and syntactic structures (Karim and Haq, 2014) as well as “macro skills” such as scanning text to locate specific information, skimming to get the gist, identifying stages of an argument, identifying examples given in support of the argument (Hamamrad 2017) reading creates problems for test designers. Berry (British Council) says depending on the purpose people tend to read in various ways: quickly, carefully, or attentively. For this reason, it is important to design Reading tests as relevant and meaningful as possible (Berry, British Council) so that tests help to evaluate learners' abilities in different language aspects and provide testers with information concerning learners' language proficiency development and level of their understanding of the taught material (Hamamrad, 2017). Hughes (2003) suggests considering intended readership, number of words, range of grammar, and topics to be taken into consideration while designing the reading test. For this reason, after the interview with the student, it was decided to choose 4 out of 25 mini texts that have somehow academic content as the reader uses English mostly for academic purposes, longer texts were chosen as it was suggested by Qu (2016) “college students are expected to master more than 5000 words and be able to read 500-or more-word passages.” Also, it was looked through the grammar range of the mini-texts and the choice was given to the ones with a bit complex grammar. The vocabulary range was left as it was taking into consideration the intended readers who are from A1-B2+ according to the CEFR descriptor (2001). Analyzing the list of the input text source designed by Alderson et.al., (2006) that can be used for modifying reading texts for the reading test, one can conclude that the modified version of the test does not use authentic texts given in the list and the texts from the test were designed for pedagogical purposes. It is suggested to choose a passage that has not previously been used in a reading comprehension test (**Module six Assessment task**). However, considering that changing text would lead to a test creation rather than test modification the texts were left as they were.

Norris (2000) says that it is essential to establish what jobs language tests should do and select the appropriate language test tool to have the job done. Being summative by its nature the purpose of the modified test was to find out to what extent the learners achieved objectives set in the curriculum. The test items were developed based on the objectives set for the strategy used in the syllabus. Table 1 shows the alignment of the test item, the syllabus objective, and the scoring.

No	Item in the test	Syllabus objective	Scoring
1	Multiple choice	Reading for Gist	1
2	Multiple choice	Reading for Specific information	1
3	Completing the sentence	Summarising-main idea	1
4	True/False/Not given	Understanding details	1
5	True/False/Not given	Understanding details	1
6	True/False/Not given	Understanding details	1
7	Meaning of pronoun in the passage	Finding reference	1
8	Fact/Opinion	Distinguishing fact from opinion	1
9	Fill in the gap	Vocabulary (productive reading)	0.5x2
10	Fill in the gap	Vocabulary (productive reading)	0.5x2

Table 1

Fengying, (2003) suggested having an objectives-referenced evaluation in order to develop the teaching process, and learning results and check the effectiveness of teaching. Taking into consideration the importance of the purpose of each specific test to be used the modified version was designed (Basanta 1995). In the “**Module six Assessment task**” it is suggested to consider giving candidates ‘a a good number of fresh starts’ by using a number of passages and for this reason, four passages were chosen rather than focusing on only one. As was suggested in the **Module six Assessment task** the language of the questions and the language of the instruction should be clear as the problems in this section may affect the learner’s attitude to the test. Before giving the test to the learner the modifier asked the instructor to look through the test and check if the instructions and questions are clear. After, some instructions were modified and the test was given to the volunteer student.

Practicality

According to Karim and Haq (2014) while testing reading, taking into consideration the time for processing the text is very essential for this reason I asked a friend of mine who is also a teacher to do the test with a set time and then doubled the time for the learner. Oydina got 30 minutes to do the test but she finished in 18 minutes. Another point to consider in practicality is teachers’ access to training to become qualified testers, graders, and administrators (Hamamorad, 2017).

Authenticity

As Hamamorad (2017) explains “authenticity is identifying to what extent a language test task corresponds to the features of an actual "real-life" language”. it is not problematic to find authentic text but creating an authentic task can be challenging (Basanta, 1995). Analysing to what extent the modified test is authentic it was mostly focused on task authenticity as the texts were taken from the previous test. It has to be admitted that while designing the test items the designer mostly thought about what information can be asked from the text rather than is the question is authentic and if people search for such kind of information in this type of reading.

Washback

Hamamorad (2017) claims that “The washback which is received from test results will provide the teacher with valuable information about appropriateness and effectiveness of adopted pedagogy and students' strengths and weaknesses as well as helping him/her formulate suitable

strategies to boost their performance in reading”. The results from this test showed that Oydina’s satisfaction towards the test was higher than before saying that “I feel I did good because it is what we usually do in our classes”.

Validity

According to Hamamrad (2017) the reading test would be valid if it assesses learner’s ability to “grasp the gist, understand details, find specific information, guess the meaning of lexical items in context and to draw inferences from a text”. The modified version was designed to represent this idea as well as being aligned to the syllabus and the classroom activities. This time the test was taken on a piece of paper as Hemis was not allowed to be used and this way represented the validity as well. Since students usually do readings on a piece of paper in class and the same was done with Oydina with modified test.

Reliability

According to Karim and Haq (2014) “the more discriminating the items in a test are, the more reliable the test is”. The test designer tried to follow this suggestion and designed seven different purpose items represented in the Table 1.

Conclusion

Each reading test should have a purpose and do its own job. The test designer should consider the different types of reading that the test taker will need to carry out in the target domain, reading processes, skills, and sources of knowledge to be involved while doing the test (Green, 2014). The Modified version of the test is more practical, semi-authentic, reliable, valid, and has a positive washback on the learner. It is based on the strategies provided in the curriculum however, it still needs to be developed. The test types are similar to the test items students get while preparing for the IELTS and TOEFL exams which would give them the feeling that the test is designed correctly.

Appendix

Pre-modification

1. 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.

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

cost more than other kinds of films.

are more popular among the elderly than among the young.

should be banned altogether.

are less damaging to young people than films of violent crime.

2. The famous Tower of London was built as a fortress by William the Conqueror. Early in the Middle Ages the kings used it as a palace; later on it was turned into a prison, but only distinguished prisoners, including statesmen and princes, were held there. Today the Tower is a national museum, where, among other things, the jewelry of the English kings and queens is on display.

It is obvious from the passage that the functions of the Tower of London _____.

were all established by William the Conqueror.

have always been controlled by the kings.

have varied greatly over the centuries.

are all of a military nature.

3. Every summer many people, girls and women as well as boys and men, try to swim from England to France or from France to England. The distance at the nearest points is only about twenty miles, but because of the strong currents the distance that must be swum is usually twice as far. The first man to succeed in swimming across the Channel was Captain Webb, an Englishman. This was in August 1875. He landed in France 21 hours 45 minutes after entering the water at Dover. Since then, there have been many successful swims and the time has been shortened. One French swimmer crossed in 11 hours and 5 minutes.

Swimming the Channel is not as easy as it might seem _____.

as the distance between the two countries is far too much.

and it always takes more or less 20 hours.

and only two people have managed to do it so far.

So, few people even try to swim it.

Post-Modification

Passage 1

Every summer many people, girls and women as well as boys and men, try to swim from England to France or from France to England. The distance at the nearest points is only about twenty miles, but because of the strong currents the distance that must be swum is usually twice as far. The first man to succeed in swimming across the Channel was Captain Webb, an Englishman. This was in August 1875. He landed in France 21 hours 45 minutes after entering the water at Dover. Since then, there have been many successful swims and the time has been shortened. One French swimmer crossed in 11 hours and 5 minutes.

Q.1-2 Read the passage 1 and answer the questions 1-2

1. What can be the title for the passage?
 - A. Crossing the channel
 - B. Uncrossable channel
 - C. Annual Event in Europe
 - D. English and French swimmers
2. How wide is the channel at the adjacent area?
 - A. 40 miles
 - B. 20 miles
 - C. twice as far
 - D. 60 miles

Q.3 Read the passage and continue the sentence. Write down no more than 5 words.

Swimming the Channel is not as easy as it might seem _____.

Passage 2

A new discovery of a dinosaur fossil in Antarctica has confirmed the idea that dinosaurs lived not only in the Northern Hemisphere but in the Southern Hemisphere as well. Up until this discovery, scientists had found dinosaur remains on every continent except Antarctica. This new discovery now confirms the idea that dinosaurs were distributed worldwide. If the dinosaur fossil is shown to be related to other dinosaurs of the same period in South America, it will also support the idea that South America and Antarctica were once linked together.

(Q4,5,6) Read the passage 2 and choose one that suits most according to the passage.

True/False/Not given

3. Dinosaurs lived only in the Northern Hemisphere of the Earth. _____
4. Dinosaur fossils found in every continent besides Antarctica. _____
5. The dinosaurs found in South and North America are from the same period. _____

Passage 3

Most people were not impressed when in 1913, the Daily Mail newspaper offered 10,000 pounds to the first pilot to fly across the Atlantic in under 72 hours. The majority of scientists, even said it could not be done. Certainly, the problems involved were many and far

ranging. Obviously, the design of the aero plane was of great importance, but so too were the skill and courage of the pilot and the navigator; weather conditions also had to be taken into consideration. A very few enthusiasts thought it might be possible ten years later. They were wrong. A pilot received the prize just six years later.

Q.7. What does the word 'they' (line 6) refer to in the passage 3?

- A. Scientists B. enthusiasts C. problems D. pilot and navigator

REFERENCE

- Basanta, C.P. (1995). Coming to Grips with Progress Testing: Some Guidelines for Its Design. *English Teaching Forum*, 33 (3), pp 37-40.
- Berry, V. Introducing Language Assessment: Transcript – Assessing Reading. Retrieved from www.britishcouncil.org/exam/aptis/research/assessment-literacy
- Brown, D.H. (2010). *Language Assessment: Principles and Classroom Practices* (2nd Edition). White Plains, NY: Longman Pearson. ISBN: 9780138149314
- Common European Framework of Reference (2001). <https://www.coe.int/en/web/commoneuropean-framework-reference-languages>
- Fengying, M. (2003). Motivating Students by modifying evaluation methods. *English Teaching Forum*, pp 38-41.
- Green, A. (2014). *Exploring language Assessment and Testing*. NY. Routledge.
- Hamamorad, A.M, 2017. Reading Skill and Test Design Principles: Designing a Diagnostic Reading Comprehension Test for Kurdish EFL learners in the University Level. *Journal of Education and Practice*, 8(30).
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- J. Charles Alderson, Neus Figueras, Henk Kuijper , Guenter Nold , Sauli Takala & Claire Tardieu (2006) Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference: The Experience of The Dutch CEFR Construct Project, *Language Assessment Quarterly: An International Journal*, 3:1, 3-30, DOI: 10.1207/s15434311laq0301_2
- Joint Task Force on Assessment of the International Reading Association and the National Council of Teachers of English. (2010). *Standards for the Assessment of Reading and Writing*.
- Karim, Sh, and Haq, N. (2014). The Process of Developing an Academic Reading Test and Evaluating Its Authenticity. *Journal of Language Teaching and Research*, 5 (2), pp. 471-475.
- Norris, J.M. (2000). Purposeful Language Assessment: Selecting the Right Alternative Test *English Teaching Forum*. 38 (1), pp.41-45.
- Module six Assessment task. Retrieved from: <https://www.birmingham.ac.uk/Documents/college-artslaw/cels/essays/testing/testing/RoseTesting.pdf>
- Oguz, A. Test Master. Baskent Education Publication. Retrieved from: <https://staff.ttiame.uz/storage/users/247/books/GnGozbEwKKuzJeFZwlP44jd4P7XPAoUUz9IMqJEX.pdf>
- Qu, Y. (2016). A Probe into Test Design of Reading Comprehension in College English. 6th International Conference on Management, Education, Information and Control.
- Regulation № 3069 on Assessment system at Higher Educational Establishments 2018 Retrieved from <https://lex.uz/docs/3916793>
- Uzbekistan State World Language University. Curriculum for Reading and Writing (2020). № BD -5111400-2.04