



MODIFYING A TEST (BY INTEGRATING TECHNOLOGY) TO MEET THE REQUIREMENT OF ASSESSMENT AND STUDENTS` NEEDS

Temurbek Saidnazarov

Tashkent State University of Oriental Studies

Abstract: *Currently, there are a variety of tasks and assignments for testing students, but not all of them are up to standard. This paper examines a test including a couple of items from different angles and evaluates how effective and demanding it is to make modifications. My research question is: what changes should be made to tests to assess students' knowledge, and how important is this? I conducted this research by observing students doing the sample tests. As a result of my observations, it has been found that tests are not always appropriate for all types of learners, so it can be recommended that teachers of language should take into account the certain principles of testing while altering exercises and tests.*

Key words: *practicality, washback, validity, authenticity, technology integration.*

Introduction

Assessment is the most influential part of education, in other words, assessment leads teaching process. Assessment is considered as an ongoing procedure including a broad range of methodological techniques (Brown, 2004). One of the techniques of the assessment is a test, according to Bachman (1990), a test should measure the test-taker`s performance with the help of a certain set of rules and procedures. A learner`s ability and knowledge can be identified and assessed by testing in a particular domain. In this project, I am meaning to investigate a placement test discovering numerous critical points, and then I will discuss the modified version of the original test providing with a couple of reasonable rationales for the modifications.

Learner assessment profile

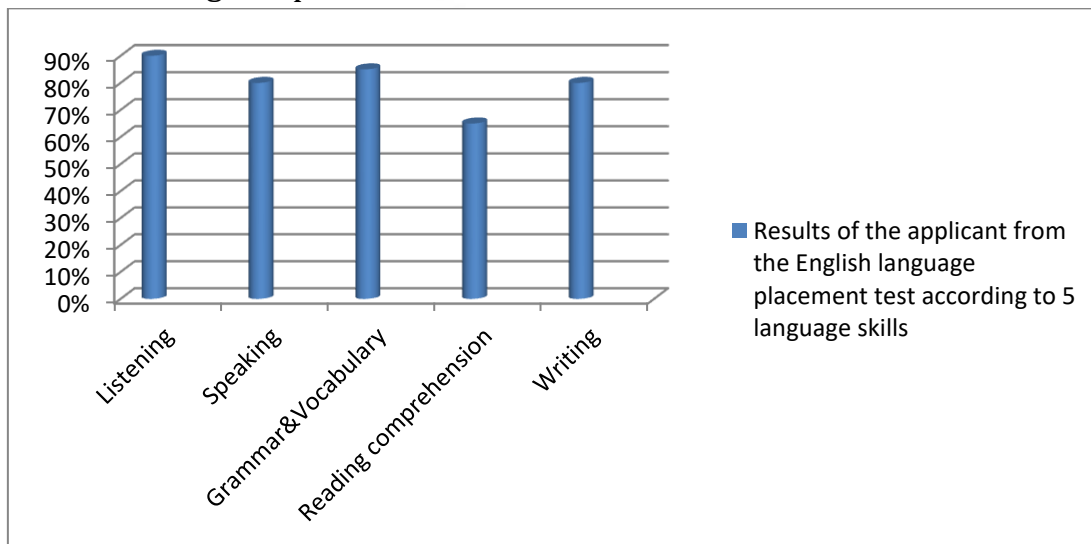
The original placement test has recently been taken by a fourth grade student, who is a ten-year-old boy. The student used to go to public school before, but he decided to enroll for a non-governmental school named "Vosiq International School". Prior to accepting the application of the pupil, the private school had required the applicant to pass an exam, where the pupil had to do a number of placement tests ranging from listening to reading tests. His first language is Uzbek, and he has been studying English language at a state-run school for three years; during his education at the public school, he participated in the international English language competition organized by Global Hippo Association in 2020. He was at Hippo 2 which is specialized for the students born in 2008 and younger (the participants compete according to Common European Framework References scaling from A1 to A2). The student was awarded with a bronze medal in this competition. Addressing his assessment





guidelines at the public school, his grades have been excellent all the time. From his achievements, it is apparent that his level is between A1 and A2 in English language.

Despite his outstanding results, Vosiq International School demanded him to take a placement test which was intended to examine the student's listening, writing, reading, grammar and vocabulary as well as speaking skills. In order to pass the exam successfully, each applicant must fulfill the school requirement namely the student's overall result must not be less than 75% from the skill-based English exam. The main reason why the private school administered the test is that the school has three types of fourth grade classes with diverse language levels and competence, and it is crucial to place the new student in an appropriate class considering his level of English. As far as his results from English language placement test, except his result in reading comprehension, all other skills are fairly good, which can be observed in the bar graph below. In this chart his results have been scaled in percentages, it is apparent that he exceeded the required percentage (75%) in terms of nearly all tests. However, the reading comprehension test taken by the applicant was not so perfect to measure the test-taker's reading competence.



In my opinion, the student had better be provided with another reading test; nevertheless, the student has been placed in the class which is made of proficient learners merely.

It is advisable for the school to administer the placement test again with better reading comprehension test taking into consideration several implications of the exam results.

According to Brown (2004), the placement test helps not only to place a learner into a particular suitable class or level, but also it is quite beneficial in receiving diagnostic data on the test-taker's performance. In consequence of a proper placement, the educators will be informed and instructed how to teach and guide the student appropriately.





Critique of an existing language test

General description of the existing test

The placement test (look at appendix 1) taken by the learner was created by an English teacher of Vosiq International School. In reality, I am not going to describe all tests based on evaluating different language skills of the test-taker. Only the reading comprehension test is due to be under discussion. If truth be told, the test administrator made the test using the passages from the 4th grade students' course books which are in use at this private school merely. The test creator just assembled the questions for the existing text. The test consists of two types of questions some of which need to be marked as true or false, which represents multiple choice tests, while the others demand open-ended answers, that is, gap filling technique.

Detailed description of the test

This multiple choice reading test (look at appendix 1) includes 10 items only, since the paragraphs of the text are brief and they do not provide the reader with so much information.

In terms of readability, factually, a reliable test depends on a couple of factors such as test administration, the scoring, the student, and the test itself (Fulcher & Davidson, 2007). Addressing the test reliability, the original test is reliable, on the one hand; on account of the type of the test. Multiple choice tests have usually got carefully designed structure, and the test-taker do not have any necessity for special creativity, he or she just has to select one of the alternatives. Brown (2004) states that multiple choice test decreases the unreliability as it can offer precise alternative options for responding to questions.

On the other hand, this test is lack of reliability, Firstly, as it has 3 items (look at appendix 1) that require open-ended responses, namely subjective tests. According to Brown(2010), in subjective tests the rater bias might cause the test unreliability, when the teacher is expected to determine the responds as correct or incorrect.

Secondly, the existing test must be completed within 25 minutes which may lead to test-taker's anxiety. Brown and Abeywickrama (2010) claims regarding timed tests that those who tend to perform badly on a test with a time constraint, might feel the sense of discrimination concerning the timed tests.

In order to check the validity of the original test, it is necessarily important to measure how useful and appropriate the test is for the learner. Gronlund (2003) explains the validity connecting the purpose of the assessment with meaningfulness, appropriateness and usefulness of the test.

The administration of the original reading test meets the requirements of reliability, because the condition where the test administered is quite preferable and comfortable for the 4th grade test-taker. In addition the exam room was light enough and free of noise pollution and other distracters. Moreover, the chairs and desks were placed specially for 10-year-old children.





For better clarification and analysis of the validity of the original test, there are some special evidences, for example:

-Content-Related Evidence

According to Hughes (2003) and Mousavi (2009), a test must measure exactly what it aims to measure. In the original placement test (appendix 1), the item 9 (*Ten centuries ago the height of the water clock in China was 1000 centimeters. True False*) does not measure the learner's reading competence, instead it was designed for challenging the learner with math, in other words, calculating the time and height measurements.

-Construct-Related Evidence

Construct validity refers to conducting a test that encompasses various components of what is being measured (Brown 2004). The original test does not meet the standards of construct-related evidence, since this existing test cannot analyze all careful reading operations that students are expected to carry out. For example, this test does not require students to make pragmatic inferences, interpret topic sentences or identify the writer's intention and the like.

-Face Validity

According to Mousavi (2009), face validity can be seen if the students consider the assessment to be fair, relevant and beneficial for enhancing learning. But the existing reading test seems to be short of face validity due to the fact that the direction for the items from 1 to 3 (*Questions 1-3. Read the text carefully, and write one or two words for each question. **The words you are going to write should be the elements of the types of the clocks/watches.***) is not crystal clear and too long for 4th grade students; plus, this reading text is not familiar to the test-taker, because it was made according to the course book of the private school unlike the public school materials. The level of the reading passages is higher as it has some complex grammar elements (for instance: *In China about a thousand years ago a man called Su Sung built a huge water clock, which was 10 meters high.*)

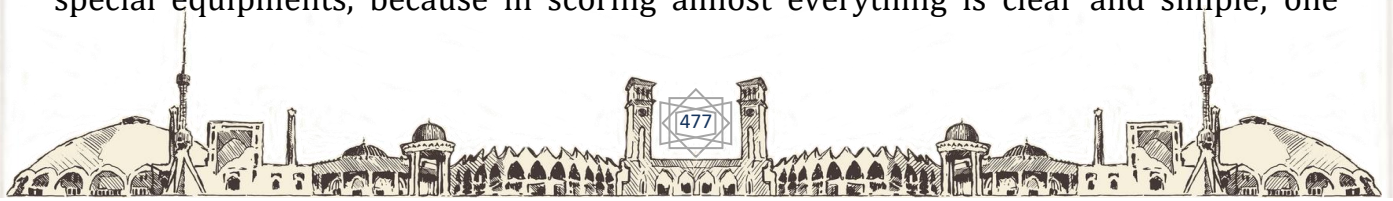
Practicality of the original test

Mousavi (2009) states that test practicality can be affected by a number of factors such as costs, amount of time for test taker and test administrator, clear scoring and reporting the results easily.

From the existing test, it is apparent that the test does not exceed budgetary limits at all, because to provide one test-taker with the test, one sheet of paper is needed, therefore, it can be regarded as practical in terms of expenses. But whenever we have to give 100 or more students the same test in paper, it will be hard to call this paper-based exam practical.

In view of time constraints for the completion of the test is 25 minutes, which is adequate for 10 questions with short text.

For designing and scoring, the test administrator is not required much time or special equipments, because in scoring almost everything is clear and simple; one





question makes for 10% of the total score (100%). Why almost everything owing to the questions from 1-3 (appendix1). In these items, students may write different words that can have relevancy or not; in this circumstance test checker probably face some trouble in grading the student's responds honestly, which refers to impracticality of the test.

Authenticity of the test

Authenticity shows the degree of correlation between the features of the provided task and the elements of the target language (Bachman and Palmer, 2015). The original test (appendix 1) contains both authentic and non-authentic parts; for instance, first and second paragraphs are about ancient clocks namely water clocks, the learner is 10 years old, so it is unbelievable that the boy is aware of this kind of devices. Even in his real life he may never see or use them. However, the rest of the text is about modern watches and clocks, which is familiar to the learner which makes the test authentic.

Washback

Hughes (2003) defines washback as effects that the test usually brings to teaching and learning.

As to the washback of the existing test, there are more negative impacts to be taken from this test. As a matter of fact that a test should give learners feedback to increase their expertise, but this test is quite summative and only indicates the test-taker's result.

The modified version of the chosen test
(<https://ru.padlet.com/temurbekbk/i21ttr7bmhgdhqmj>)

Meeting the learner's needs is of great importance, in consequence, I modified the placement reading test in conformity with the test-taker's age, level, culture and interests (look at appendix 2). Bachman (1990) stated four feasible roots of test bias: "cultural background, background knowledge, cognitive characteristics, native language/ethnicity/age and gender". Test designers are advised to avoid these kinds of test bias as possible.

Jones and Jones (2000) confirm that student's needs include feeling the sense of importance and safe in the educational environment, comprehending learning goals, possessing time to integrate learning, fully realizing the learning process, as well as getting feedback. If their needs are met in class, their motivation to acquire knowledge will increase significantly, indeed.

The first item that I altered is the structure of one statement in the original text. The sentence constructed in the original test (*In China about a thousand years ago a man called Su Sung built a huge water clock, which was 10 meters high*) was grammatically complex for the 10-year-old student's level, so that I separated the long sentence into two simple statements, which can be observed in the second paragraph of the modified text (look at appendix 2).





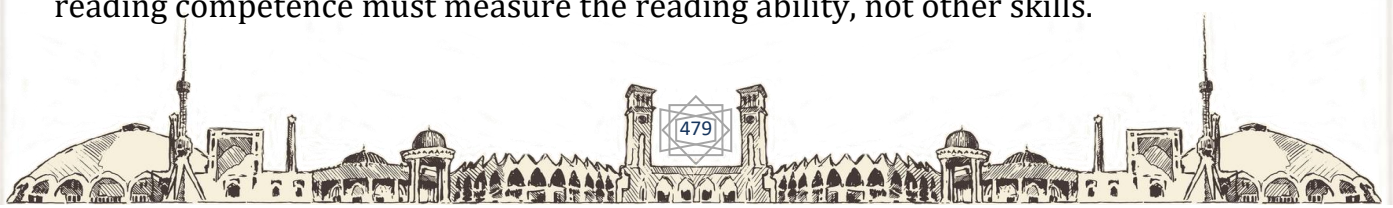
Second and the most influential modification includes the addition of relevant pictures for each question (look at appendix 2). The first reason is related to student-related reliability of the test, for example, when the learner looks through the original reading test, he felt anxiety and unawareness of the topic and context. According to Mousavi (2009), the psychological impact of the context should be taken into account, as most learners tend to become nervous and anxious in the test administration context. Additionally, by integrating the visuals into questions we can increase the practicality of the test in terms of time constraints. Because while the learner is looking through the questions, the pictures attached deliver the meaning of the sentences faster and clearly, as a result, the student can finish the test earlier. As Brown (2004) stated, a test-taker must be sure to complete the task within the limited time making the test practical.

Thirdly, I altered the initial three questions in the original test from gap filling to matching strategy (look at appendix 1,2). This modification supports the test reliability of the test reducing the rater bias; in that, in the original test the student had a hesitation to write the most appropriate words to the three items, plus, the teacher had difficulty in determining correct and incorrect responds due to the numerous alternatives. In the modified version (appendix 2), the student matched the three letters to the numbers given with types of clocks. Moreover, the change including the sort of the initial three items boosts the consequential validity of the test. Mainly because when the test-taker is going to reply to the three questions in the modified test, he needs to read the whole text and comprehend what is being described, which gives information about the learner's input process to the teacher eventually.

Next, I have transformed the paper-based reading test into internet-based test using Padlet.com, because the test-taker is studying online at the moment. Integration of modern technologies into our life enabled us interact with each other, notwithstanding the distance (Zhao, 2003). Information, Communication, and Technology (ICT) stimulates daily classroom instructional process. In addition to preparing students for the present computerized era, teachers are regarded as the main characters in utilizing ICT in their classes. This is owing to the significance of ICT in providing dynamic and dynamic teaching-learning environment (Arnseth & Hatlevik, 2012).

The next modification occurred to improve the authenticity of the original test. I changed the 4th question in the original test (the 5th question in the modified version), seeing that the item was not contextual and precise. According to Brown (2004), an authentic test must be meaningful and contextualized.

The last but not least item I altered is the question 9 in the existing test (question 8 in the modified test, appendix 2), because the question was not valid enough. The item did not seem to examine student's reading comprehension, instead, it was based on some mathematical calculations. As Brown (2004) mentioned, a valid test of reading competence must measure the reading ability, not other skills.





The final change deals with the sequence of the items, Brown (2004) named this situation as “episodic items”, according to the paragraphs. In reality, the original test did not offer sequential organization of the questions, which makes the test less authentic. Therefore, I decided to replace the items considering the sequence of events from the context.

I have just administered the modified version of the existing test with the same student online. I sent the link of the test and set the time. When he finished I was surprised a lot, because his result from this test was 95%, and he completed it in 15 minutes, 10 minutes earlier. Then I asked for his opinions concerning the test; he mentioned that as soon as looking at pictures under the questions, he got rid of anxiety and fear in doing this test.

Conclusion

It is my confession that I have revised the all articles from Language testing within a couple of days, thanks to this project. So far I did not use to pay attention to the purpose, meaning, principles of the testing or assessment, when I was supposed to create a test. I realized that every single thing can affect the quality of the assessment; if the test that we designed is adequately authentic, reliable, practical and influential, moreover, if it meets the learner`s needs, the assessment or the test can take the teaching process to the highest of the success.

Appendix 1 (Original Reading Test)

Read the text below, and do the tests below within 25 minutes, in total.

Some people used water to tell the time

This is a water clock from Greece. Water flowed through a small hole in the higher jar and slowly filled the lower jar. The amount of water in the jar told you the time.

In China about a thousand years ago a man called Su Sung built a huge water clock, which was 10 meters high. Water turned a big wheel and every hour, bells and drums played.

Clocks and watches

In the 16th and 17th centuries people began to use small metal wheels to make clocks and watches. At first, they only had one **hand** to show the hours. Later they had another hand to show the hours. Today we have many kinds of clocks and watches to help us to tell the time.

Questions 1-3. Read the text carefully, and write one or two words for each question. The words you are going to write should be the elements of the types of the clocks/watches.

- 1) Water clock in Greece: _____ . (10%)
- 2) Water clock in China: _____ . (10%)
- 3) Today`s watches and clocks: _____ . (10%)

Questions 4-9. Read the text above, and decide these statements whether they are true or false.





4) Su Sung was a Chinese inventor. True False (10%)

5) In Greece people looked at the water in the jar to tell the time. True False (10%)

6) In 16th and 17th centuries, clocks and watches were made of small metal wheels. True False (10%)

7) The word "hand" in the third paragraph means one of the person's organs. True False (10%)

8) Before two hands of the clock would show the time, but after some time the clock had one hand left. (10%)

9) Ten centuries ago the height of the water clock in China was 1000 centimeters. True False (10%)

10) The special tools of Chinese water clock were a big wheel, bells and guitar. True False (10%)

In total: 100%

Answer Key of the Original Test

1) (small) hole/ (higher) jar/ (lower) jar/ water

2) water/ (big) wheel/ bells/ drums

3) (metal) wheels/ hand

4) TRUE

5) TRUE

6) TRUE

7) FALSE

8) FALSE

9) TRUE

10) FALSE

Appendix 2 (Modified Reading Test) Time limit is 25 minutes.

Read the text carefully, and do the tests below. Each correct answer makes for 10% of the total score, 100%.





Some people used water to tell the time

This is a water clock from Greece. Water flowed through a small hole in the higher jar and slowly filled the lower jar. The amount of water in the jar told you the time.

In China about a thousand years ago a man called Su Sung built a huge water clock. It was 10 meters high. Water turned a big wheel and every hour, bells and drums played.

Clocks and watches

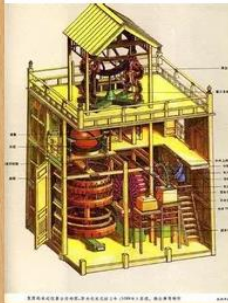
In the 16th and 17th centuries people began to use small metal wheels to make clocks and watches. At first, they only had one hand to show the hours. Later they had another hand to show the hours. Today we have many kinds of clocks and watches to help us to tell the time.

Questions 1-3

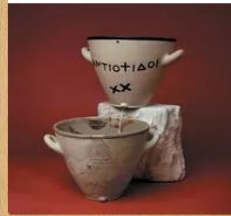
Read the text carefully and match the types of clocks below with the pictures. Write the correct letters for each number above the three pictures

- A) Today's clock
- B) Water clock from Greece
- C) Water clock in China

1) _____



2) _____



3) _____



Questions 4-10

Read the text given at the top, and highlight the statements as true or false.

- 4) In Greece people looked at the water in the jar to tell the time.
True False



Questions 4-10

- 5) Su Sung is builder of a huge water clock.
True False



Questions 4-10

- 6) The special tools of the Chinese water clock were a big wheel, bells and guitar.
True False



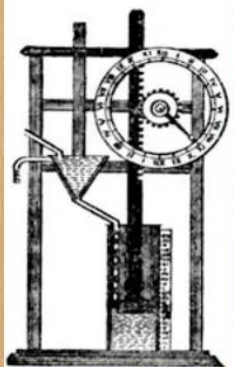
Questions 4-10

- 7) In 16th and 17th centuries, clocks and watches were made of small metal wheels.
True False



Questions 4-10

- 8) The Chinese had a 10-metre water clock many years ago
True False



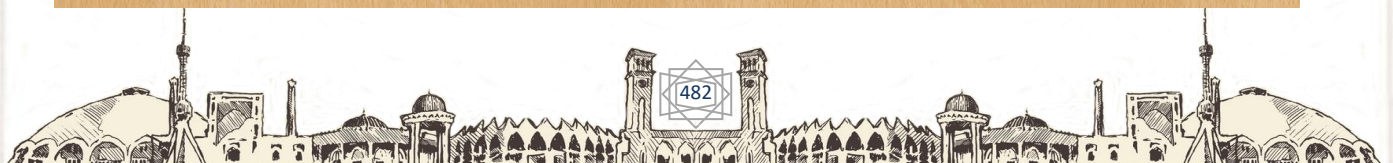
Questions 4-10

- 9) The word "hand" in the third paragraph means one of the person's organs.
True False



Questions 4-10

- 10) Before two hands of the clock would show the time, but after some time the clock had one hand left.
True False





Answer Key of the Modified Test

- 1)D
- 2)B
- 3)A
- 4)TRUE
- 5)TRUE
- 6)FALSE
- 7)TRUE
- 8)TRUE
- 9)FALSE
- 10)FALSE

Appendix 4 (Power Point Slides)

Plan

- 1) **learner assessment profile**
 - a) -details of the learner
 - b) -educational setting
- 2) **Critique of the existing language test**
- 3) **Modified version of the test**



Learner Profile

- age: 10 years old
- grade: 4th
- L1: Uzbek language
- English background: 3 years at a public school
- L2: English
- level of English: A1+ and A2
- Hippo





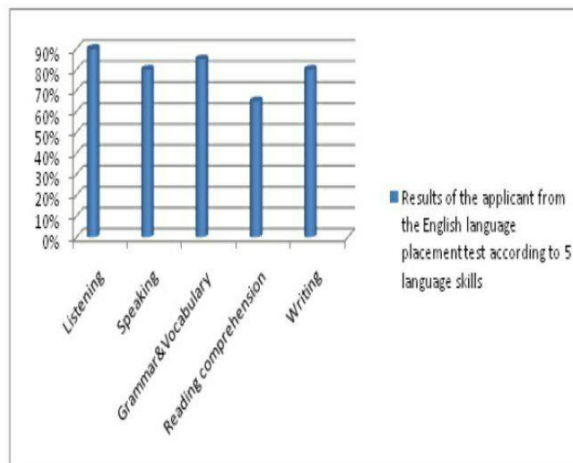
Educational Setting

Used to study in a public school
3 year at a public school
Vosiq International School
Private
In Tashkent
English - main/aspect



The Chosen Test

-placement test
-multiple choice
-gap filling
-for A1+ students
-at high stake



1) Critique of the existing language test

- Multiple Choice (true/false)/Gap filling
- Reliability - **Brown (2004/2010) multiple choice + /open ended-/ timed.**
- Validity** - Gronlund (2003)- the purpose of the assessment with meaningfulness, appropriateness and usefulness
- Content-Related Evidence** - Hughes (2003) and Mousavi (2009), a test must measure exactly what it aims to measure





Critique of the existing language test

Authenticity - (Bachman and Palmer, 2015)

Water clock-non-authentic modern clocks-authentic

Practicality - Mausavi (2009) -administration-time-money

Face validity - the instruction is not clear

existing test

VS

modified test

-paper based

-internet-based

-multiple choice/gap filling

-multiple choice/matching

-long sentences

-sentences with pictures

-grammatically complex

-grammatically simple

-cultural unawareness

-culture awareness

-lack of validity

-valid enough

REFERENCE:

Arnseth, H. C., & Hatlevik, O. E. (2010). Challenges in aligning pedagogical practices and pupils' competencies with the information society's demands. *Cases on Interactive Technology Environments and Transnational Collaboration*, 266-280.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2015). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.





Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson/Longman.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Abingdon, England: Routledge.

Gronlund, N. E. (2003). *Assessment of student achievement*. Boston: Allyn and Bacon.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Jones, V., and L. Jones. 2000. *Comprehensive classroom management: Creating communities of support and solving problems*. 6th ed. Needham Heights, Mass.: Allyn and Bacon.

Mousavi, S. A. (2009). *Language Testing*. Maleny: EContent Management Pty.

Zhao, Y. (2003). A Comprehensive Review of Research on Technology Uses in Language Education. *The CALICO Journal*, 21(1), pp.7-27.

