

An EFL placement test and its use in a private high school.

By Elsa Fernanda Gonzalez

UNIVERSITY OF BIRMINGHAM

Centre for English Language Studies

Open Distance Learning Master of Arts in Teaching English as a Foreign/Second Language (MA TEFL / TESL)

Module 6: Testing

Assignment Task # TS/11/01

September 2011

Student ID # 1050678

Critically evaluate an established language test with which you are familiar. This may be either a test used in your local teaching situation or a large-scale language test such as those provided by Cambridge ESOL (www.cambridgeesol.org). The assignment should include critical discussion of the following.

- Test background including the test purpose, clientele, test context and the underlying theory of the test.
- Test design, content and procedure.
- Overall test characteristics including the validity, reliability and practicality of the test. The assignment should be supported by reference to the relevant academic literature.

CONTENTS

1	INTRODUCTION	4
2	THEORETICAL BACKGROUND	4
2.1	Testing and its Concept.....	4
2.2	Test Specifications.....	4
2.3	Placement Tests.....	5
2.4	Important Test Characteristics.....	5
2.4.1	Validity.....	5
2.4.1.1	Construct Validity.....	6
2.4.1.2	Content Validity.....	6
2.4.1.3	Criterion Validity.....	7
2.4.1.4	Face Validity.....	7
2.4.2	Washback.....	8
2.4.3	Reliability.....	8
2.4.3.1	Rater Reliability.....	9
2.4.3.2	Test Creation Reliability.....	9
2.4.3.3	Test Administration Reliability.....	10
2.4.4	Practicality.....	10
2.5	Reliability and Validity Interaction.....	10
3	AN ANALYSIS OF A PLACEMENT TEST	10
3.1	Test Purpose and Target Students.....	10
3.2	Test context.....	11
3.3	Test Design and Content.....	12
3.4	Testing and Rating Procedures.....	13
3.5	Validity.....	14
3.5.1	Content Validity.....	14
3.5.2	Criterion Validity.....	14
3.5.3	Face Validity.....	15
3.6	Reliability.....	15
3.7	Practicality.....	16
4	CONCLUSION	16
5	REFERENCES	18
6	APPENDIXES	19
6.1	Appendix A: Placement Test.....	19
6.2	Appendix B: Writing Section Scoring Rubric.....	25

1. INTRODUCTION

When the use of tests is hindering students' development or interfering with the teaching goals, tests are not fulfilling their purpose. Therefore, it is necessary to follow specific language curriculum objectives and use the most suitable test types to achieve those goals. As pointed out by Hughs (2003: 8), before moving on to the creation and use of a test, teachers or testers need to have a clear purpose for testing.

This paper has the purpose of analyzing a placement test currently used at my work place. The test is used as a measurement and placement tool of newly enrolled high school students. I begin this paper by presenting a brief discussion of the relevant theoretical background of testing and important test characteristics such as validity, reliability and practicality. Then, I describe the test under discussion and the context in which it is used. Finally, I provide an analysis of its creation, use and characteristics.

2. THEORETICAL BACKGROUND

2.1 Testing and its concept.

According to Hedge (2002) testing is an opportunity learners are provided to demonstrate their knowledge about language and their use of specific language skills. For authors such as Hughs (2003) tests are tools that can aid teachers in assessing their students' development in the classroom. Testing is unfortunately necessary at some point in a language curriculum and in the teaching context. Although sometimes avoided, testing is needed. *'It is difficult to imagine, for example, British and American universities accepting students from overseas without some knowledge of their proficiency in English'* (ibid: 4).

2.2 Test specifications.

According to Brown (2004) and Hughs (2003), test specifications are an outline of test characteristics that have the purpose of informing test takers, test validators or public in general (Alderson, 2000) test objectives, content and other important traits. Information such as *content, structure, timing, medium or channel, techniques to be used, criterial*

levels of performance, and scoring procedures (Hughes, 20003: 59) are considered and described in this document. Its purpose is to be a guide in the test creation process and providing test takers an opportunity to understand what to expect in the test.

2.3 Placement tests

Placement tests have the purpose of assigning students a specific level of language ability within the curriculum they wish to be incorporated to (Hughes, 2003; Harmer, 2007; Brown, 1994). It is common for book editorials to include placement tests as part of their textbook pack. However, tests should be created according to institution's specific needs (Hughes, 2003). In other words, testing is a feature of language teaching that may best work if it is set to fulfill a specific context and the test taker needs. According to the Educational Testing Service (2007), placement tests carry advantages such as increased student learning with their incorporation to their corresponding proficiency level, reduced student and faculty frustration and increased student retention. Specific test principles should be cared for to assure test success such as validity, reliability and practicality. These items are further discussed in the next section.

2.4. Important Test Characteristics.

2.4.1 Validity

A test is said to be valid if it measures what it was created to measure (Brown, 2004; Hughes, 2003). Validity is the *'extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment'* (Gronland, 1998:226 cited in Brown, 2004: 22). Tests are created to fulfill a purpose, which frequently is to measure skill ability; therefore a test is valid if it actually measures skill ability (Hughes, 2003: 26). According to some researchers, total validity in a test is difficult to obtain (Spolsky, 1999).

Over the years, the concept of validity has changed. In the 70s, validity was approached in large-scale research while it was unimportant for classroom purposes. In this decade three important types of validation were identified: Content, construct and criterion validity (ibid). However, the 90s lead to a new vision of validity. It started focusing not only on a test's content but also considered its use and consequences. Construct validity

is considered central while content and criterion validity are used as evidence to constructs (ibid). Researchers such as Chapelle (1999) have attempted to describe validity. She considers that for validation to be measured hypotheses or assumptions regarding tests need to be formulated and further analyzed.

2.4.1.1 Construct Validity.

Brown (2004: 25) defines a construct as a theory, hypothesis or model that explains observed phenomena. Therefore, proficiency and communicative competence are considered linguistic constructs. A test construct needs to be validated and described by underlying theory that explains the nature of the test (Hughes, 2003). *'If all of the items in a test were meant to measure specified abilities, then, without evidence that they were actually measuring those abilities, the construct validity of the whole test would be in question'* (ibid: 31). In other words, if a test has the purpose of measuring written skills and raters consider organization, content, range, accuracy and coherence as the main traits to assess then it is necessary to identify the theoretical background that identifies these 5 aspects to be considered components of proficient written skills. The whole point of constructing a test is to measure students' specific skills based on theoretical constructs (Hughes, 2003). Nowadays, construct validity has been considered a pinning point to which content and criterion-referenced validity contribute to evidence.

2.4.1.2 Content Validity

A test's content is valid when it demands from test takers to perform the language ability the test intends to measure (Brown, 2004) and when the content of the test is a valid sample of the intended measured language (Hughes, 2003). For instance, the validity of a test that is intended to measure writing skills and includes gap-filling exercises is at stake. *'Content validity is established by showing that test items are a sample of a universe that the investigator is interested in'* (Fulcher & Davidson, 2007). A valid test includes the relevant structures to the purpose of the test (Hughes, 2003: 26) that depend on the test specifications that the institution pointed out. Test specifications are discussed in part 2.2 of this paper.

Authors have pointed out the importance of understanding direct and indirect testing as a

way of seeking for validity. While direct tests encourage the test taker to actually perform the task that is being measured, an indirect test uses tasks to measure receptive and productive skills that lie beneath the activity (Brown, 2004; Harmer, 2007: 382). Therefore, a classroom test that includes direct testing might be more inclined to obtain content validity (Brown, 2004).

Another system for obtaining validity is by comparing the language knowledge (test results) a learner has with a test taker's performance (candidate's actual use of language). This is known as Baker's model of direct vs. indirect and performance referenced vs. system-referenced testing (Baker, 1989: 11 cited in Owen, 2010)

2.4.1.3 Criterion Validity.

This type of validity evidence refers to the extent the established criteria against which test results are measured are actually reached (ibid). The test is administered and results obtained are compared to previously established criteria that test the same linguistic features such as a commercially produced test or observed subsequent behavior (Fulcher & Davidson, 2007 & Brown, 2004). Criterion validity falls into two categories, a) concurrent validity and b) predictive validity (ibid). Concurrent validity is applied at the moment of students' performance and test results are measured to pre-established skills or functions that candidate's need to meet. Predictive validity is the degree to which a test can predict learner's future performance (Hughes, 2003). Establishing criterion validity can be difficult especially when, in the case of predictive validity, a clear picture of what future success entails or describing the exact level of agreement between criterion and tests is unclear (ibid). The calculation of a validity coefficient (ibid) can provide a numerical measurement of the agreement between tests and its criterion to meet. A numerical coefficient of 1 points out a high degree of agreement while a coefficient of 0 depicts total lack of agreement (ibid). *'Whether or not a particular level of agreement is regarded as satisfactory will depend upon the purpose of the test and the importance of the decisions that are made on the basis of it'* (ibid: 28).

2.4.1.4 Face Validity

Rather than a scientific and research factor, face validity refers to the opinion of test

takers, instructors, administrators or any other person involved in the testing process (Brown, 2004; Hughs, 2003). It refers to the extent to which a test seems, on the surface, adequate and appears to measure what it was made to measure (Mousavi, 2002 cited in Brown, 2004). It is an issue that researchers believe may represent psychological consequences on learners such as anxiety or confidence (Brown, 2004). For instance, a test that is meant to test pronunciation but does not carry a speaking component may cause in test takers a lack of confidence in the results obtained.

It seems that finding true and crystal clear validity is a complicated task. Validating procedures can be time consuming and cost ineffective. However, establishing test validity is considered an important responsibility of any test user (Chapelle, 1999).

2.4.2 Washback

Also known as consequential validity (Brown, 2004) or backwash (Hughs 2003), this form of validity evidence focuses on the consequences of testing on teaching and learning (ibid). Test consequences can be negative or positive and can include impact of test preparation, effects on learners, social interpretation of tests, effect on student motivation, subsequent performance in a course, study habits, or student attitude towards courses (ibid). If test preparation takes up all teaching objectives instead of teaching language skills, then the washback effect tends to be negative (“teaching to the test”, ibid). However, when weaknesses and strengths are identified and represent an opportunity for improvement, washback can become a positive effect.

2.4.3 Reliability.

Test reliability describes a test that is dependable and consistent (Brown, 2004). It is not possible to trust test results without reliability (Davies, 1999). A reliable test, when administered to different students on different occasions, describes similar results. Just as validity, reliability is a condition that can be represented numerically with a reliability coefficient. Its specific calculation procedure is not discussed for it goes beyond the scope of this paper. It can also be measured by comparing the actual test with an external and highly dependable test (Davies, 1999). However, for teaching and classroom

purposes, reliability can be analyzed considering factors such as students, raters, test creation and test administration (Brown, 2004).

2.4.3.1 Rater Reliability

Raters may also represent a dangerous factor that can jeopardize test reliability. A situation in which 100 writing test papers need to be graded by the same rater may cause fatigue or human error and are exposed to reliability flaws (Brown, 2004; Hughs, 2003). *Inter-rater reliability* describes a situation in which two or more scorers obtain distinct results on the same test due to the use of different scoring criteria, inexperience, or preconceived opinions (ibid). On the other hand, *intra-rater reliability* refers to the lack of accurate scoring of a single teacher result of default scoring criteria, fatigue, carelessness, or bias towards particular students. It is of major importance for test creation and test scoring to have a reliable balance among them. If test reliability is not reached then reliability in scoring is not either (Hughs, 2003).

Researchers have attempted to find ways of minimizing unreliability in tests. For instance, Uysal's (2010) paper critically reviews the scoring procedure of the writing component of the English International Testing Service (IELTS). IELTS authorities implement strategies such as a multiple rating system, training sessions for raters and pairing up single examiners with a senior examiner to rate a paper. Uysal comments that to avoid reliability issues in the grading of writing tasks, multiple raters should be used permanently and reliability coefficients calculated regularly.

2.4.3.2 Test Creation Reliability.

Test nature needs to be cared for when revising test reliability. For instance, timed tests may cause in students stress and fatigue that may interfere with their performance. If it's too long or too short, tests may not represent a reliable measurement of students' true ability (Brown, 2004). Hughs (2003) considers that reliability can be assured if a) enough items of a specific ability are included, b) items discriminate between learners' weaknesses and strengths, c) tests avoid providing too much freedom in their items, d) items are unambiguous, e) instructions are clear and explicit, f) tests are well laid-out and legible, g) learners are familiar with format and testing techniques, h) use objective scoring instead of subjective as much as possible, i) scorers are trained, j) detailed scoring

keys are provided, k) candidates are given number identification instead of using their names, and finally l) use of multiple scoring on each test.

2.4.3.3 Test Administration Reliability.

Adequate lighting, photocopy quality, audio perception, desk and chair condition, temperature in room, or room distracters are part of administration reliability (Brown, 2004). Rooms in which traffic noise does not allow listening to audio may interfere with listening skill reliability.

2.4.4 Practicality.

A practical test is one that is not excessively expensive, it maintains appropriate timing, its administration does not represent difficulties and its scoring and evaluation procedures are specific and time efficient (Brown, 2004). For instance, a test that can only be answered on a computer but the school has only two computers available for a class of 20 students is impractical. Practicality refers to test administration and test scoring (Brown, 2004) and it becomes critical when practicality needs to be considered above content validity. For instance, time constraints may deprive tests of the speaking component jeopardizing content validity. Money issues need to be considered. That is, a reliable test is practical if its creation and implementation processes correspond to the school budget and administrative details (ibid).

2.5 Reliability and Validity Interaction.

Reliability is fundamental to find validity. That is, a valid test needs to provide adequate measurements that are reliable. However, a test that is reliable is not necessarily valid (Hughes, 2003). Over emphasis on reliability can risk losing validity. The more items included in a test may increase reliability but risk validity (Davies, 1999). Therefore, it is important for language instructors and test creators to balance the gains and losses of both validity and reliability aspects (Hughes, 2003).

3. AN ANALYSIS OF A PLACEMENT TEST

3.1 Test purpose and target students.

This paper has the purpose of analyzing a placement test (Appendix A) used to enroll new students in the EFL program of a private Mexican high school.

It is used at the beginning of each new semester by the language department to measure new students' English language abilities and place them in the correct level of the English language within the curriculum.

Candidates are teenagers whose ages range from 15-18 years old. On most occasions, they have some English learning background that allows them to be placed in high-beginner or low-intermediate levels.

This placement test is of major importance for learner's language development. Therefore, I consider indispensable to point out test specifications that should be followed. In section 2.2 of this paper, it is mentioned that Hughs (2003) and Brown (2004) consider test specifications describe test structure and what it contains (Alderson, 1999). However, this test is not accompanied by a test specification document.

In terms of format, the test consists of 100 points divided among grammar and vocabulary items, a reading task and a writing assignment. Grammar/vocabulary includes 27 fill-in-the blank items, 11 matching items, 23 multiple-choice and 5 short answers. The reading task requires students to find specific information and fill in a chart (9 points) while the writing task requires students to write a short description of minimum 100 words about themselves (25 points). This test is included as Appendix A.

3.2 Test context

The English curriculum is composed of six proficiency levels and has the objective of developing communicative competence in the four skills at a B2 level according to the Common European Framework. Learners take part of a daily two-hour class. Due to administrative issues, levels need to have a least 10 students to be offered.

The placement test is used to measure student's English proficiency and place them in their corresponding level of English. At the beginning of each semester, new enrollments are gathered in a single room and are given the placement test to answer. They are not given a specific time limit to answer. However, they are expected to complete it before the two-hour English class is over. If students have doubts during the test regarding

instructions or answering procedures, test administrators are permitted to help. The use of dictionaries or mobile devices during the test is prohibited.

Tests are rated by a group of teachers, which include the English coordinator who is responsible for test design, administration and grading process. Other raters are teachers who are part of the English teaching staff and have participated in the administration process. Teacher raters are provided with an answer key for the grammar/vocabulary and reading sections and analytic rubric is used to rate the writing section (Appendix B). As pointed out in section 2.4.3.1 of this paper, intra-rater reliability may become flawed when rater's scoring criteria is obscured by fatigue or other personal characteristics. Therefore, the use of a single rubric may minimize reliability risks.

Test results are announced the following day. Candidates are placed in their levels according to the number of points obtained as an overall score. The maximum number of points obtained is 100 while the minimum is 0. Scores ranging from 75 to 100 points are placed in an intermediate level, 74 to 55 in low intermediate and 54 to 21 are placed in beginner and 20-0 in introductory levels. During a personal interview, the English language coordinator mentioned that administrative issues and limitations in addition to test results are part of the placement criteria. Factors such as students' learning background (if known), overall language proficiency, number of semester the student is coursing, classroom capacity, and teacher availability are part of her considerations when making placement decisions.

3.3 Test design and Content.

As mentioned in section 2.3 of this paper, a placement test seeks to incorporate students into a specific level of language ability in the curriculum (Hughes, 2003; Harmer, 2007; Brown, 1994). It is context specific and is made to suit the specific needs of the language program (Hughes, 2003).

The placement test under analysis includes three sections: grammar and vocabulary (66 points), reading (9 points) and writing (25 points). Neither listening nor speaking is tested. Although this test follows the purpose of a placement test, it does not correspond to the specific purpose of the language curriculum of the school: to achieve language competence from levels A1 to B2 of the CEFR. I consider since this test only includes

two language skills and an extended grammar/vocabulary section, it does not focus on communicative competence.

Test items are tested directly and indirectly. As pointed out in section 2.4.1 direct testing encourages the test taker to actually perform the task that is measured (Brown, 2004). For instance on exercise V, test takers are required to choose an option from four answer choices that test their ability to use quantifiers, time expressions, comparatives and verb tenses. Question 1 states,

1. Are there _____ soft drinks in the fridge?
a) any b) a c) an) no

In this case, the question directly tests the use of quantifiers but it is also indirectly testing the ability to use the grammar structure in a specific context. On the other hand, this item has two possible answers, which may jeopardize content reliability.

Exercise VI requires learners to discriminate information in a passage and fill in a chart with specific information. Although it directly tests reading skills, I consider items are insufficient to provide a true picture of learners' real reading abilities.

The writing section requires candidates to produce a short description of them. This task is basic and might seem easy but in this case written ability is tested directly by having students produce written text.

3.4 Testing and Rating Procedures

Testing procedures are equally followed on every occasion the test is administered. Candidates are taken to a classroom that is big enough for the number of students (if necessary two rooms are used). Listening skills are not tested therefore sound is not considered. However, temperature and lighting conditions are to be suitable. Each student is given a legible copy of the test and is required to answer directly on the test. No answer sheets are required. I believe this allows students to be more in control of their answers and avoid making mistakes when transferring their answers to an answer sheet.

Grading procedures are carried out immediately after all candidates have concluded their test. The English coordinator and the teaching staff (5 teachers in total) are responsible

for the rating of tests. Tests are divided equally among teachers to be rated. Although, multiple raters participate in the scoring procedures one rater grades a single test. I consider this method can be improved and it is discussed in section 3.6 of this paper.

3.5 Validity

Section 2.4.1 states a test is valid *'... if it measures accurately what it is intended to measure'* (Hughes, 2003: 26). To analyze whether or not the validity of a test is met, test specifications need to be described. As mentioned in section 2.2, test specifications aim at informing test takers, valutors or public in general test characteristics (Alderson, 2000). However, a test specification document was not found. There is no description of what is to be tested or the specific objective of the test. Moreover, candidates are not informed of test content, grading procedures or test administration specifications. Therefore, establishing validity and reliability for this test can become an obscured task.

3.5.1 Content Validity

Content validity assures that test items correspond to the skills and abilities the test is meant to test. Therefore, a classroom test that includes direct testing might be inclined to obtain content validity (Brown, 2004). I believe this placement test lacks full content validity for it only includes two language skills: reading and writing. This test over relies on grammar and vocabulary instead of the four language skills. I consider grammar/vocabulary important but if the goal is communicative competence (as pointed out by the head of the language department), a listening and speaking component should be added. Additionally, I consider more items for the reading and writing section would provide more opportunities of skill measurement.

3.5.2 Criterion Validity

The importance of criterion validity lays in the comparison of students' test results against a reliable and highly dependable source of reference such as another commercial test or observed behavior (Fulcher & Davidson, 2007; Brown, 2004). Predictive validity, as pointed out in section 2.4.1.3, allows students' future success with the language to be predicted. Therefore, this validity evidence is crucial for placement tests. In this case, the

placement test procedures followed do not include a comparison process to assure criterion validity. Therefore, it is my belief that this process needs to be implemented for predictive validity to assure successful student placement. On the other hand, other student characteristics (students' language learning background) are considered for placement, which I believe, may jeopardize the predictive validity of the test.

3.5.3 Face Validity.

As pointed out in section 2.4.1.4, face validity refers to the extent to which a test seems, on the surface, adequate and appears to measure what it was made to measure (Mousavi, 2002 cited in Brown, 2004). Students' and teachers' opinions are crucial. Having to answer 100 items may cause students to feel fatigued. For teachers, grading a long test may appear to be exhausting. These issues may have an effect on students overall performance and the rating process. I would consider shortening the test to avoid these face validity issues and include more specific items. On the other hand, the lack of listening and speaking sections may cause discomfort in test takers or test administrators. Most learners seek above all to have good speaking skills. Therefore they may feel the test cannot be trusted if this section is not included.

3.6 Reliability

Once answered, placement tests are divided among teachers to lessen time investment. A single test is rated by one teacher, which I believe gives over-reliance on the rater's subjective opinion especially during the assessment of writing tasks. Even though different raters participate, a single teacher rates each test. Therefore, I believe multiple raters for a single test will avoid flawed intra-rater reliability. As described by Uysal (2010), multiple raters should be used permanently, especially working with written tests. Therefore, having at least two teachers rate a test could help avoid using the subjective opinion of a single rater.

In terms of administration reliability, raters grade students' answers on the test and are aware of the name of each candidate. I believe the use of answer sheets and code names for each candidate will help rating procedures reliability by avoiding raters become aware of the person they are grading and subjective opinions be made about candidate.

As mentioned in section 2.4.3.2, test reliability can be assured if a) enough items of a specific ability are included, b) items discriminate between learners' weaknesses and strengths, c) tests avoid providing too much freedom in their items, d) items are unambiguous, and e) instructions are clear and explicit (Hughes, 2003). I consider this test does not provide students any type of freedom in items nor it includes exercises that are unclear or unambiguous. However, it does not include enough items to fully test writing skills.

3.7 Practicality

I consider this to be a practical test even though it includes 100 points to grade. Grammar/vocabulary and reading are closed-responses in which only one answer is possible. Therefore, its grading process is quick. Additionally, the test does not include a speaking or listening component. If a speaking section were included, as with the writing section, the use of a rubric could aid in minimizing rater subjectivity and time investment. Avoiding speaking tasks may facilitate the implementation process and raise practicality but it puts at risk test validity. Additionally, the avoidance of the listening section makes the use of audio equipment not necessary facilitating the work of the test administrator. Copies for every student assures they have the conditions needed to answer a reliable test. In terms of budget and administrative conditions, I consider this test is practical for its implementation does not exceed the available budget at the institution. The only economical investment the school makes is in photocopies. The teaching staff is not paid extra for their rating. It is part of their activities at this school.

4. CONCLUSION

Throughout this paper, I attempted to analyze a placement test I am familiar with. Although I consider it a good and reliable exam, I found aspects such as content and rater reliability could be handled with more care. The use of multiple raters and more skill-aimed content items in the test are some examples of test improvements I may suggest. The responsibility of validity and reliability is usually left to language managers or coordinators to solve. Especially in schools in which teachers are not responsible of test design. However, writing this paper has brought to my attention how important it is for

classroom teachers to become involved in the validity and reliability processes because the validity and reliability of a test is the responsibility of all those who use it (Hughes, 2003).

5. REFERENCES

- Alderson, C. (2000) Test Specifications. In Fulcher, G. and Thrasher, R. *Language Testing Videos*. In association with ILTA, available: <http://languagetesting.info.s>
- Allison, D. (1999) **Language Testing and Evaluation: An Introductory Course**. Singapore University Press: Singapore
- Brown, H.D. (1994) **Principles of Language Learning and Teaching**. Third Edition. Prentice Hall: New Jersey, USA.
- Brown, H.D. (2004) **Language Assessment Principles and Classroom Practices**. Pearson Education Longman: White Plains, New York, USA.
- Chapelle, C. (1999) Validity in Language Assessment. **Annual Review of Applied Linguistics**. 19, 254-272
- Davies, A. (1999) Reliability. In Fulcher, G. and Thrasher, R. *Language Testing Videos*. In association with ILTA, available: <http://languagetesting.info.s>
- Educational Testing Service (2007) Developing an ESL Placement Test for Community College Students. Available www.ets.org/Media/Research/pdf/conf_achgap_cc_young.pdf
- Fulcher, F. & Davidson, F. (2007) **Language Testing and Assessment: An Advanced Resource Book**. Routledge: Oxon, United Kingdom.
- Harmer, J. (2007) **The Practice of English Language Teaching**. Fourth Edition. Pearson Longman:
- Hedge, T. (2002) **Teaching and Learning in the Language Classroom**. Oxford University Press: Oxford U.K.
- Hughs, A. (2003) **Testing for Language Teachers**. Cambridge University Press: Cambridge, United Kingdom.
- Owen, C.; Miller, N.; Rees, J. & Crompton, P. (2011) "What makes a test any good?" *Testing CELS*, Birmingham.
- Spolsky, B. (1999) Validity. In Fulcher, G. and Thrasher, R. *Language Testing Videos*. In association with ILTA, available: <http://languagetesting.info>.
- Uysal, H.H. (2010) A Critical Review of the IELTS Writing Exam. **ELT Journal** 64/3, 314-320.

6. APPENDIXES

6.1 Appendix A: Placement test

SENIOR HIGH ENGLISH PLACEMENT TEST

Name: _____ Former school: _____

GRAMMAR

I. Choose from the box the correct option. Write the letter on the line.

a) he has	b) have	c) Do	d)her	e) he studies
f) she's	g) loves	h) is she	i) does she	j) she is

A: (1) _____ your sister?

B: Yes, (2) _____. Her name is Johanna. (3) _____ a high school teacher. Her husband is an engineer. (4) _____ a software company. They (5) _____ two kids, a boy and a girl.

A: Really? What (6) _____ the children do?

B: The boy's name is Alex and (7) _____ medicine at Harvard University.

A: What about (8) _____ daughter?

B: She is in high school and (9) _____ painting.

A: Wow! That's interesting. (10) _____ take painting lessons?

B: I really don't know.

II. Fill in the blanks by writing in the parentheses the correct number.

CONVERSATION A

A: Excuse me. (11) _____ furniture store near here?

B: Yes. Villarreal (12) _____ great furniture. They (13) _____ a sale right now.

A: And a bank?

() it's

() has

() there aren't

() is there

() are having

() I moved

B: No, (14) _____ any banks around here.

() did

A: Are you new here?

() I'd like

B: Yes, (15) _____ in today.

() did

() went

() was

CONVERSATION B

A: (16) _____ your friend arrive from Paris?

B: Yes, she (17) _____. Her flight

(18) _____ late.

A: (19) _____ to meet her.

B: She (20) _____ to visit her mother because

(21) _____ her birthday. We can visit her tomorrow.

III. Complete the conversation. Use the correct forms of the words in parentheses.

A 22 _____ (do / this weekend)?

B 23 _____ (go / rock concert).

24 _____ (would like / come)?

A No, thanks.

25 _____ (see / Mary).

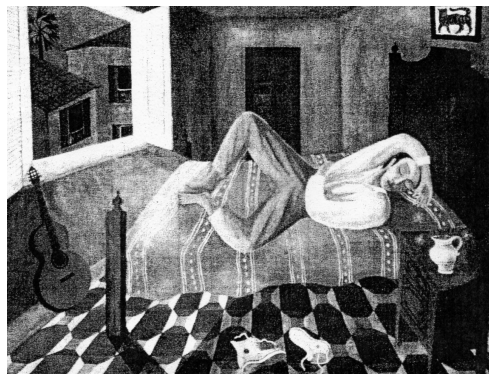
B 26 _____ (good time)!

A Thanks.

IV. Look at the painting and complete its description. Use the words in the box. Write the corresponding letters on the lines.

a. bed b. center c. chest d. door e. houses f. on g. left h. man i. pants j. picture k. under l. right m. shirt n. shoes o. wall p. window q. next to

This is a painting of a bedroom. A (27) _____ is sleeping (28) _____ the bed in the (29) _____ of the room. He's wearing a white (30) _____ and gray (31) _____. He isn't wearing any (32) _____. On the (33) _____ of the picture, there's an open (34) _____. It's daytime, and you can see two (35) _____ outside. There's a guitar (36) _____ the window. There's a (37) _____ in the wall at the back of the room. On the (38) _____ of the picture, there's a small (39) _____ of drawers next to the (40) _____. There are some sneakers (41) _____ the bed, and a (42) _____ of an animal on the (43) _____.



V. Complete the sentences with the most appropriate option. Circle your response.

44) Are there ___ soft drinks in the fridge?

- a. any b. a c. an d. no

45) Yes, there are ___ in the refrigerator. But you ___ drink soft drinks all day.

- a. them, should b. some, don't c. any, aren't d. some, shouldn't

46) **A:** How ___ bottles of water do you drink? **B:** Two or three bottles daily.

- a. many b. much c. any d. drinks

47) You ___ drink water. It's better for you.

- a. don't b. have to c. isn't d. shouldn't

48) I like this sweater. It is _____ than the other one.

- a. pretty b. prettier c. more pretty d. the prettiest

49) This is the _____ exam of the year. I have to study a lot.

- a. important b. more important c. most important d. the most important

50) I have to go to bed early. My flight _____ at 6:00 am. tomorrow.

- a. left b. is leaving c. will leave d. had left

51) _____ cake did you eat?

- a. How any b. How much c. How many d. What

52) **A:** How long have you studied English? **B:** _____ I was in kindergarten.

- a. for b. ago c. since d. yet

53) Really? I have taken lessons only _____ a year.

- a. for b. ago c. since d. yet

54) I can't choose between the red dress and the black one. _____ are beautiful!

- a. It b. both c. neither d. some

55) George and Mary did the project all by _____.

- a. itself b. himself c. herself d. themselves

- 56) The weather is beautiful, ___?
 a. doesn't it b. hasn't it c. isn't it d. is it
- 57) A: I don't like math... B: ___ do I!
 a. so b. neither c. too d. no
- 58) If you ___ an aspirin, your headache ___ go away.
 a. will take, go b. take, will go c. will take, will go d. take, would
- 59) You forgot to call me, ___?
 a. didn't you b. did you c. do you d. don't you
- 60) I am very tired. I ___ for hours.
 a. have been studying b. am studying c. will study d. had studied
- 61) The garden was dead because it ___ dry all summer.
 a. is dry b. had been c. would be d. will be
- 62) Harry Potter and the Goblet of Fire ___ by J.K. Rowling
 a. written b. writes c. is written d. was written
- 63) Yesterday I asked Bill _____ to the party.
 a. are you coming b. if he was coming c. if he come d. if he comes
- 64) I ___ this book three times now. It is fantastic!
 a. am reading b. read c. have read d. have been reading
- 65) The calculator will turn _____ off when it is not being used.
 a. itself b. herself c. themselves d. ourselves
- 66) If I ___ a million dollar, I ___ around the world.
 a. have, will travel b. had, will travel c. had, would travel d. have, would travel

READING

VI. Read the text about famous places in China. Fill in the chart with the correct information

TOP ATTRACTIONS OF CHINA

The Great Wall of China is one of the great wonders of the world. It runs 6700 kilometers across deserts, grasslands, and mountains. The construction of the wall began in the 3rd century A.D. and was rebuilt in the 15th and 16th centuries. It was originally built in separate pieces to protect China from invaders in 214 B.C.

Emperor Qin Shi Huang unified the country and he joined the walls together into one huge wall. It took ten years to finish. The Great Wall is a monument to the building skills and to the wisdom and patience of the Chinese people.

The Museum of Qin Terra Cotta Warriors at Xi'an is a sight not to be missed by any visitor to China. Life size clay figures of 7,000 soldiers, horses and chariots stand in battle formation. These famous figures were discovered by chance in 1974. A group of workers found ancient pottery when they were digging a well. The site caught the attention of archaeologists immediately. The warriors are from the tomb of Emperor Qin Shi Huang. When he came to power between 221-206 B.C., at the age of 13, he began work on his tomb. It took many years to complete.

In the Center of Beijing, to the north of Tiananmen Square, lies the Forbidden City, now known as the Imperial Palace Museum. It was the home to the emperors of the Ming and Qing Dynasties. It is the world's largest Palace, covering an area of 74 hectares, and it includes 9,999 rooms in 800 buildings. Construction of the palace began in 1406 and ended in 1420. It took 200,000 workers and 100,000 artisans to build the Forbidden City. The Palace Museum contains many rare treasures and curiosities, such as gates, halls, and an imperial garden. The palace is now one of the most popular tourist attractions in the world.

Famous Attraction	Where is it?	What is it?	Important Characteristics (Mention at least 2)
The Great Wall			
Terra Cotta Warriors			
The Forbidden City			

WRITING

VIII. Write a short paragraph (100-150 words) in which you describe yourself. You may include the following information: (25 points in total)

- a) General information like your name, age, address, etc.
- b) Physical aspects like hair or eye color and height.
- c) Family information.
- d) Your likes, dislikes and hobbies.
- e) Future plans and goals.

6.2 Appendix B Writing Section Scoring Rubric

ANALYTIC SCORING

Adapted from: 1) Jacobs et al, 1981 2) Cohen, 1994

CRITERIA	DESCRIPTION	Score
<i>Content</i>	Do students approach assigned topic or task appropriately? Are main ideas stated clearly? Is it substantive enough? Are ideas or opinions stated and developed accurately? Are details stated and supported appropriately? Can students communicate their ideas?	
<i>Organization and Cohesion</i>	Is content well organized? Are ideas logically sequenced? Does student write fluently? Is it cohesive?	
<i>Vocabulary</i>	Is vocabulary used adequate? Was there effective word choice? Is vocabulary repeated frequently? Are idioms appropriately used? Is there adequate register? Is misuse of vocabulary obscuring meaning?	
<i>Grammar Accuracy</i>	Are there effective constructions? Are tense, number, articles, pronouns, and other functions appropriately used? Are accuracy mistakes obscuring meaning?	
<i>Mechanics and Spelling</i>	Does student use paragraphing? Are punctuation, capitalization, and spelling accurately used? Is handwriting legible? Is meaning obscured?	