

Evaluative commentary on a language test

Trang W. NGUYEN

Abstract

Language tests occupy a central role in the cycle of language learning and teaching. The major issue in any language assessment is whether the test measures what it intends to measure. This paper argues that when designing and implementing language tests, it is crucial for developers of language tests and for teachers to clearly understand the function of each assessment activity and what that activity essentially measures. This test review will employ Bachman and Palmer's (1996) model of language assessment as an organizing framework to examine the credibility and usefulness of a classroom-based language test. It first describes the test then discusses aspects of the test in relations to the framework's qualities of reliability, construct validity, authenticity, interactiveness, impact, and practicality. The analysis identifies a number of problematic features seen to contain in the test under review. The paper concludes with reflections on the value of using a language assessment model as an organizing framework to examine the credibility and usefulness of a classroom-based language test.

INTRODUCTION

Assessment plays an important role in the cycle of learning and teaching. In language education, assessment can be used to measure achievement, gauge proficiency, appraise learners' strengths and weaknesses, or to determine appropriate placement of the learner in a course or level of study (Hughes 2003: 8). In the broader context of society, moreover, assessment can also serve as a "disciplinary tool" and as a form of social and "political control", such as in the implementation and maintenance of vocational training, immigration and citizenship policies of different governments (Kunnan 2005: 780; McNamara 1998; McNamara 2000: 68-72; Bachman & Palmer 1996: 34; Bachman 1995: 291). In light of the important and diverse roles that assessment plays in language education and in view of its wider impact on society, it is inevitable that much debate surrounds the design, implementation and use of assessment.

The main aim of language assessment is to provide a *measure* based on which a person's language competency and ability can be appraised (Hughes 2003: 50; Bachman and Palmer 1996:

23). Essentially, for assessment to be recognized as being valid and credible, the assessment must be able to demonstrate that the resulting scores accurately measure what the assessment intended to measure, and that these scores sustain meaning beyond the testing situation and relate specifically to particular language competencies (Hughes 2003: 26; Alderson 1981: 55). In language assessment, it is crucial that the actual testing evaluates learners according to the targeted language skills and that it does so through relevant text and types of tasks that relate to the targeted language competencies (Bachman and Palmer 1996: 21; Chapelle 1998: 50-51).

Bachman and Palmer (1996: 17) propose that authenticity, reliability, construct validity, interactiveness, and practicality are the qualities by which it may be possible to measure and come to an understanding of the appropriateness and usefulness of any given piece of assessment. They suggest that in order to ascertain whether a given assessment item or activity is valid, it is crucial to examine the “usefulness” of that assessment and explore whether the assessment fulfills its intention and does what it set out to do in ways that reflect and encompass the language learning within a given context. As such, Bachman and Palmer (1996: 23) construe the authenticity and validity of the language test to be the extent to which the facets of the assessment activities correlate to the language tasks, which they label as “target language use” (TLU) tasks. What is deemed important, then, is that assessment is created with a clear purpose, for it to be located within a certain context that corresponds to “real-life” (Bachman 1995: 301-305) language use beyond the testing situation, and for the assessment to be designed and aimed towards a specific group of learners or participants.

This test review will employ Bachman and Palmer’s (1996) model of language assessment as an organizing framework to examine the credibility and usefulness of a classroom-based language test. It first describes the test then discusses aspects of the test in relations to the framework’s qualities of reliability, construct validity, authenticity, interactiveness, impact and practicality.

TEST PURPOSE

The test under review is a second-year mid-term examination administered to students at a two-year junior college in Japan as part of the institution’s English Language Program. The purpose of the test is to measure learner knowledge of English as a foreign language during the course of an academic term. As such, it is an ‘achievement’ test. There are four such tests given in a year for the one subject. The test makes inferences about the reading abilities, writing abilities and vocabulary knowledge of the students undertaking the course.

LANGUAGE LEARNING CONTEXT

The English Language Program incorporates the skills development of reading, writing, speaking, and listening as part of the course aims. There are six levels ranging from Beginner, Lower intermediate to Advanced in this Program. The particular test under review encompasses some of the content of the lower-intermediate level syllabus, and reflects the types of writing and reading tasks contained in the textbooks used as part of the teaching-learning activities in the course. In addition to this test, students also participate in other forms of assessment such as writing book reports, undertaking speaking tests and poster presentations with class participation, classroom English usage, attendance and effort, making up the rest of the term total grade. The final mark assigned for the subject impacts on students' future academic and employment opportunities.

TEST ADMINISTRATION

The tests are developed by the teachers of English Language program and administered during the college examination periods in the classrooms where the subject is taught. Students are given 45 minutes to complete the test. The test is collected at the end of the examination session by the supervising teacher and marked by the teacher. The scores are tallied and the test returned to students for feedback and discussion in subsequent lessons.

TEST DESCRIPTION

The test is a paper-and-pencil test (see Appendix). It has eight parts, A-H. Part A requires students to write five sentences using five given words. The expectation is that students will write grammatically correct sentences and incorporate the specified words into the sentences. Part B incorporates statements relating to English expressions and the concepts of *facts* and *opinions*, requiring students to write *True* or *False* next to each statement. Part C requires students to match vocabulary to their meanings. Part D asks students to match five underlined words that appear in five sample sentences to their synonyms. Part E comprises of five sample sentences and related questions about verbs, nouns, and adjectives to which students answer *Yes* or *No* to indicate their responses. Part F requires students to recognize correct English verb-noun expressions relating to sporting activities and to tick the correct answers. Part G involves reading comprehension whereby students read a passage of approximately 200 words in length and complete seven statements about the text selecting the correct answer in a multiple-choice format. Part H asks students to choose from three given topics and to write a short essay 'using good paragraph structure and complete sentences'. A word-limit for the essay is not specified.

Scores and scoring procedures

Each part of the test specifies the maximum points awarded for that particular activity with the essay (Part H) given the heaviest weighting (20 points) followed by the reading comprehension exercise (Part G, 14 points). Part A which requires students to write sentences for given words is attributed a total of 10 points. The remaining sections score one point each for every correct answer. The overall weighting of the test is 20% of the entire term grade.

TEST QUALITIES

Bachman and Palmer (1996: 17-19) propose that the elements by which it may be possible to measure and ascertain the appropriateness and usefulness of any given piece of assessment includes the qualities of reliability, construct validity, authenticity, interactiveness, impact and practicality. They suggest that in order to ascertain whether a given assessment item or activity is valid, it is crucial to examine the “usefulness” of that assessment and explore whether the assessment fulfills its intention and does what it set out to do in ways that reflect and encompass the language learning within a given context. It is essential for assessment to be created with a clear purpose, for it to be located within a certain context that corresponds to “real-life” language use beyond the testing situation, and for the assessment to be designed and aimed towards a specific group of learners or participants (Bachman 1995: 301-305).

Reliability

Reliability refers to “consistency of measurement” (Bachman and Palmer 1996: 19-21), which suggests that a valid piece of assessment is one that yields perpetually precise measurements across different forms, times, and with different individual test markers. Only when test scores are consistent can the assessment conducted offer information about the skills that are being evaluated and measured (Hughes 2003: 50; Bachman and Palmer 1996: 23).

The test under review consists of a number of sections in which scores can be assigned for clear-cut answers, which serves to strengthen the test’s reliability. Parts B to G of the test consist of definite answers for which no matter when, where or who marks the test, it is assured that the scores would remain consistent to reflect the test responses. The test, in general, contains item types that allow for great ease of scoring such as multiple-choice (Parts F, G), and choosing *Yes/No* (Part E) or *True/False* (Part B) answers. However, the weakness of such items in evaluating student competency is that they may not truly measure students’ knowledge or language ability for the responses recorded may have been a matter of chance in students making correct guesses rather than being a demonstration of answers given based on reflection, understanding and ability

(Hughes 2003: 76-77).

Another problem with such item types in testing procedures is that they imply an “absolute truth” in language comprehension while the actual ‘real-life’ process of reading and comprehending text, such as in Part G of the test, involves interpretation of multiple meanings and forms (Shohamy 2001: 24). Hence, while the ease of scoring may strengthen the test’s reliability, it serves to equally challenge the test’s validity.

Where reliability could be directly threatened in the test relates to the marking of the sentences generated by students in Part A and the essay in Part H. Here, the marks assigned by the test marker may not be consistent throughout all the tests being graded since there is no clear scoring system outlined and what features are being looked for in expression and presentation apart from the directive for students to employ ‘good paragraph structure and complete sentences’ as stated in the given test instructions. Hence, even when the same teacher grades these works, the subjective interpretations and understandings held by the individual may potentially vary and fluctuate at different moments, resulting in a variety of marks assigned to different essays. The problem is compounded when there are different markers involved. Subjectivity in marking may threaten reliability of the test and therefore, its validity.

Nevertheless, Freedman (1993) and Shohamy (2001) point out that such writing exercises under testing conditions need to remain open and refrain from becoming too objective, constrained “by time, content and scoring rubrics, and carefully controlled raters who are led to agree on one correct answer with no deviations” (Shohamy 2001: 24). In striving to attain reliability and “consistency of measurement”, there is the potential that the validity of the test may become compromised when the subjective nature of a task is manipulated into an objective procedure and the activity of writing an essay strays from the essence of true written expression.

Variability in how the test is administered might also raise doubt as to the reliability of the test. While there is a general rule against cheating in exams across educational institutions, how strict teachers are in enforcing the rules and how they set up the conditions for testing may be inconsistent. For example, some teachers may insist on students sitting one seat apart from one another and for students to maintain absolute quietness, remaining in their seats until the end of the examination session. Other teachers, however, may not specify seating arrangements and may tolerate non-verbal communication between students, permitting them to leave the classroom immediately after their test has been completed. Such dissimilarity in testing conditions

coupled with various other factors including the amount of time permitted to complete an activity, and the level and detail of instructions given, may bring about variable learner performance in testing situations. Rea-Dickins and Gardner (2000: 236 - 237) note that these factors can potentially affect “the opportunities for learners to demonstrate their language ability” which would mean that the results arrived at are not a true reflection of learners’ abilities and knowledge, hence compromising the test’s reliability and validity.

Validity

Construct validity is recognized as pertaining to the overall concept of *validity* in assessment that also encompasses the notions of content validity and criterion-related validity (Bachman 1995: 241- 291; Chapelle 1999: 257-258; Hughes 2003: 26; Messick 1980, 1988). Chapelle (1998: 33) articulates *construct* as being “a meaningful interpretation of observed behavior”, while Bachman and Palmer (1996) perceive it as “the specific definition of an ability that provides the basis for a given test or test task”(p.21). In other words, construct refers to the responses given by participants on any given tests that enable interpretation of ability.

When speculating about the construct validity of an interpretation of a test score, Bachman and Palmer (1996: 21-23) suggest that it is vital to identify the construct definition and examine the features that the test tasks incorporate so as to be able to ascertain the degree to which the test tasks correlate to “real-life” language interaction and use. It is this relationship or congruity between the task features and the content embodied in the assessment that largely contributes to the validity of the assessment.

Examining the test under review, it may be assumed that the construct to be measured through this assessment procedure is the knowledge component of English vocabulary, grammar and that of English language use through the acts of reading and writing. The language knowledge component is tested through the tasks of distinguishing between *a fact and an opinion* and indicating whether the related statements are *True* or *False* (Part B); matching words to their meanings (Part C); matching words to their synonyms (Part D); distinguishing between verbs, adjectives and nouns with *Yes* or *No* answers; and identifying grammatically correct statements (Part F).

Mehnert (1998) reports on research findings that suggest that the fluency of language learners’ speech may be measured through the size and use of their vocabulary, while Skehan (1996) has proposed that lexical units occupy a central role in language acquisition throughout all

the stages of an individual's language growth. Although the increase of lexical knowledge is a crucial aspect of language learning, and the practice of vocabulary assessment a valid measure of language development (Read and Chapelle 2001: 3), the approach adopted in this test regards vocabulary as a distinct element of language knowledge that can be examined independently from grammatical structures, text or discourse. The testing of vocabulary here follows the more 'traditional' approach in which the design is *discrete* (Read 2000; Read and Chapelle 2001: 4 – 5) in that it seeks to inspect students' vocabulary development primarily through classifying and assessing knowledge of individual target words without a particular context to suggest how the words may be employed.

The format of the test demonstrates that vocabulary is perceived as a separate construct since assessment of the content of words is achieved through tasks such as multiple choice and word-definition matching which are relatively decontextualized item types (Read and Chapelle 2001: 5). Learners are expected to show knowledge of words in isolation and demonstrate their ability to use target vocabulary by writing sentences containing the given words. As such, the features that the test tasks incorporate do not correlate strongly to "real-life" language interaction and use, serving to compromise the test's validity.

Another construct that appears to be measured through this assessment procedure is language use through the reading of a passage of less than 300 words in length then choosing multiple-choice responses (Part G), and the writing of an essay from three given topics (Part H). It may be assumed that the content of the reading passage relates to topics that students studied during the term since the theme of environmental concerns appears as a possible topic choice for the essay-writing component in Part H of the test. An examination of the essay themes shows that the topics may be deemed to be relevant to students' lives and personal concerns. Furthermore, certain target vocabulary that appear in Part A of the test resurface as possible essay themes in Part H (such as, "What are the things you *appreciate* in your life?"), which suggests evidence of content validity.

In relation to the reading comprehension activity, Kobayashi (2002) observes that the response format in reading comprehension tests affects students' performance and their test results. She reports that cloze-tests have been shown to measure students' local understanding of the text but not their overall comprehension, whereas open-ended questions can gauge students' comprehension of the main ideas contained in the reading passage. Summary writing, on the other hand, is a "whole-text, super-macro-level skill" (Bensoussan and Kreindler 1990: 57) that can

measure students' overall understanding and general comprehension of texts. However, the response format that may be used in measuring reading comprehension in language tests depends on language proficiency and level of difficulty targeted as part of the assessment procedure.

Examining the test under review, it appears that even though Part G is meant to test reading comprehension, the multiple choice responses relate more to word-recognition rather than to text comprehension. For example, questions 32 to 37 require students to complete sentences such as:

32. *Sea turtles are* _____.

- a. *dangerous*
- b. *endangered*
- c. *danger*

The skills that students would demonstrate in completing this task is word-recognition of *endangered* being used with the words *sea turtles* in the text, then having the ability to put the lexical items together as part of the response, rather than showing understanding of the meaning of the word *endangered* in the context of *sea turtles* or other endangered species in the animal world. From a total of seven questions relating to the reading passage, only the very last question seeks to examine students' understanding of one of the main ideas in the text by asking "What is the writer's opinion?" Even here, however, students do not need to express an opinion to demonstrate their understanding of the ideas conveyed in the passage, but are required to simply choose one answer out of a possible three given responses.

Kobayashi (2002) notes that despite its wide usage as a test format for assessing reading comprehension in language tests, "the multiple choice format has a significant drawback in that test takers can guess the right answer without fully understanding the reading passage, and thus test validity is questionable"(p.197). This is especially relevant when the multiple choice format of a test offers only three choices as responses. Hence, while this particular testing format may be convenient and easy to use, it can greatly undermine the validity of an assessment procedure if its use is not appropriate. In this case, when the construct is to measure reading comprehension, but the testing format measures another skill such as word-recognition, the task cannot be recognized as "a meaningful interpretation of observed behavior"(ibid.), which serves to impair the validity of the testing activity, and therefore, the test.

AUTHENTICITY AND INTERACTIVENESS

Bachman and Palmer (1996) suggest that in order to ascertain whether a given assessment item or activity is “authentic”, it is crucial to examine the “usefulness” of that assessment and explore whether the assessment fulfills its intention and does what it set out to do in ways that reflect and encompass the language learning within a given context. Spence-Brown (2001) also conceptualizes authenticity as the extent to which the assessment process and its properties correspond to the ‘real world’ features of “language use and interaction”. She suggests that authenticity in assessment may be appraised according to the actual assessment task, the criteria and procedures, as well as the “interactiveness” between all parties involved in the assessment process. However, she makes an important point when articulating that the fundamental issue surrounding language assessment “is that tests are, by their very nature, artificial contexts for language use”(p.464). No matter how closely the assessment tasks reflect “real-life” situations or language interactions, when the context of the activity is for “assessment” purposes, it alters the way participants perceive and engage with the task and the language interaction so that these become “artificial”. Spolsky (1985) echoes this when he states that “any language test is by its very nature inauthentic, abnormal language behaviour, for the task is not to give so much as to display knowledge”(p.31).

While a language assessment item or event may *reflect* and *incorporate* real-life situations and language features, it may, however, not *replace* or *exist as* a “real-world” language event for that is not its nature nor is it its function. It remains, nonetheless, crucial for language assessment texts and activities to incorporate classroom learning, while also reflecting “real-life” settings and language use that are consistent with curricula, instruction and learning goals, for this will allow the principle of “fairness” to be operationalized in assessment and permit the learner’s language communication abilities to be gauged.

Viewed in this way, it seems that the current test is lacking in authenticity and interactiveness. For example, the act of selecting grammatically sound statements about sporting activities and ticking the correct responses (Part F) is an activity that would be inauthentic in the ‘real world’ beyond the language test. Similarly, using reading passages that are less than 300 words in length with multiple choice questions would seem to promote language comprehension that is limited to understanding single sentences rather than overall reading comprehension, which may not be appropriate to academic reading practices, which in this case would be the “real-life” language use beyond the testing situation for this particular group of learners. Weigle (2000)

rightly observes that:

Focusing on close, careful reading of short, unrelated passages to find the answers to specific detailed questions may not prepare students to cope with large amounts of reading in academic courses, to consider their own or the author's purpose in reading, or to integrate what they read with their background knowledge(p.453).

Given the criteria of an assessment item having correlation with "TLU" tasks that correspond to "real-life" language use and interaction, it would appear that the item types of the test under review do not demonstrate a high degree of validity and usefulness for the language learning purposes of this particular group of students.

Impact

Bachman and Palmer (1996) emphasize the importance of recognizing the impact tests have on individuals, institutions and the society at large because "the very acts of administering and taking a test imply certain values and goals, and have consequences" (p.30). When assessment is recognized as embodying the values and goals of a society or group of people, it seems that the idea of objectivity found in the absolute truths of *Yes/No*, *True/False* and right-and-wrong answers that dominate the test under review appear to be very limiting in vision, scope and approach (Lynch 2001). The assessment tasks incorporated may also serve to negatively impact on the students and teachers who use the test since the vocabulary and reading tasks do not seem to go beyond word-recognition, decoding word meanings and syntactic understanding, thus limiting more interactive, in-depth language use. When assessment is focused on easily testable forms of language recognition and grammar points, students may develop a lack of motivation to learn beyond what they will be assessed on.

Although the test under review is classroom-based assessment rather than a standardized language test such as the internationally accepted Test of English as a Foreign Language (TOEFL), its impact is deep and crucial to those who take part in it, nevertheless. Rea-Dickins and Gardner (2000) note that while there is tendency to regard classroom-based assessment as low-stakes, the authors argue that in certain situations "the classroom context represents major high-stakes ground. Wrong decisions have very serious implications for the individual and/or groups of pupils" (pg 238). In this case, the marks students gain on the test and the ensuing overall grade for the subject affects their future academic and employment opportunities. If a student obtains a 'fail' for the subject, s/he must complete the course requirements the following term

before being able to graduate, whereas obtaining a 'pass' grade could mean entry into another academic program, a better chance at gaining desired employment, and graduation from the college. Such direct impact of the test may have influence on students' motivations and how they approach the assessment, as would also the choice of testing methods and the item types and response formats employed as part of the overall assessment process.

Practicality

Practicality refers to the relationship between assessment development and implementation, and the resources available to fulfill these ends. Resources that are required to operationalize any given assessment include time, equipment, materials, space, and human resources (Bachman and Palmer 1996). To fulfill the requirements of practicality and fairness, the test must also be accessible geographically and financially to the test-takers, while safeguarding that the test-takers have a degree of familiarity with the equipment and the conditions (Kunnan 2004).

In this instance, the test is provided at the college location as part of the term assessment, and is administered in classrooms where students studied this same subject, which suggests that students will find the testing conditions and environment familiar and accessible. Since the test is a paper-and-pencil test, it would be feasible that this would pose no problems for students in terms familiarity and access to the equipment required to complete the test.

CONCLUSION

Since the test under review consists largely of discrete-point testing, it carries the criticisms of this approach to language testing. The emphasis placed on the use of context-independent vocabulary and grammar tasks for the ease of scoring at the expense of more in-depth and interactive language use impairs the validity of the test. However, the essay component embedded in the test serves to offer a more extensive measure of students' ability to engage in written expression that is consistent with language development in an academic setting.

While it is not possible to eradicate inconsistencies entirely in language assessment, as Bachman and Palmer (1996) have suggested, it is possible to reduce the effects of factors that compromise test validity by incorporating item types, response formats and testing methods that will enable for inferences to be made regarding individuals' language skills and abilities (Kobayashi 2002). It is essential, then, for language assessment texts and activities to incorporate classroom learning, while also reflecting "real-life" settings and language use that are consistent with curricula, instruction and learning goals, for this will contribute towards the validity of the

assessment and permit for learners' language proficiency to be gauged effectively.

In conducting this test review employing Bachman and Palmer's (1996) model of language assessment, it has been found that such models or checklists are valuable as an organizing framework to examine the credibility and usefulness of a classroom-based language test. Such checklists serve as a useful guide towards establishing basic standards or operational values in the establishment of assessment that is encompassing of language knowledge and human ability, and that precisely measures what the assessment seeks to measure. These language assessment models also demonstrate that assessment validation is a continuous process that commences when assessment is first developed and is maintained throughout the life of the testing procedure (Alderson and Banerjee 2002: 79). At the core of assessment and testing is, in the words of Alderson and Banerjee (2002: 80),

an understanding of what language is, and what it takes to learn and use language, which then becomes the basis for establishing ways of assessing people's abilities.

Purposeful, guided and well-considered assessment that incorporates a balance of principles is valuable in enhancing the cycle and experience of teaching and learning, and has the potential to impact positively the educational, social and political functions served by assessment.

REFERENCES

- Alderson, J.C. (1981). Report on the discussion on communicative language testing. In J.C Alderson and A. Hughes (Eds.), *Issues in language testing, ELT Documents 111*, London: The British Council.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Third impression. Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bensoussan, M. and Kreindler, I. (1990). Improving advanced reading comprehension in a foreign language: Summaries vs. short-answer questions. *Journal of Research in Reading*, 13, 55-68.
- Chapelle, C.A. (1998). Construct definition and validity inquiry in SLA research. In L.F. Bachman, L.F. & A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.
- Chapelle, C.A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272.
- Freedman, W.S. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment*, 1(1), 27-52.
- Hughes, A. (2003). *Testing for language teachers, second edition*. Cambridge: Cambridge University Press.
- Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, 19(2), 193-220.

-
- Kunnan, A.J. (2004). Test fairness. In M. Milanovic, C. Weir & S. Bolton (Eds.), *Europe language testing in a global context: Selected papers from the ALTE conference in Barcelona*. Cambridge: Cambridge University Press.
- Lynch, B.K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4): 351-372.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35: 1012-27.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H.I. Braun (Eds.), *Test validity*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Read, J. & Chapelle, C. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1),1-32.
- Rea-Dickins, P. & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, 17(2): 215-243.
- Shohamy, E. (2001). *The Power of Tests*. Essex: Pearson Education Limited.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38-62.
- Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing*, 8(4), 463-481.
- Weigle, S.C. (2000). Test review: the Michigan English language assessment battery (MELAB). *Language Testing*, 17(4), 449-455.

APPENDIX

English Reading and Writing Test

Student name: _____ Student number: _____ Test total score _____ / 70

A) Write sentences using the words below. (10 points)

1. priorities

2. patience

3. gradually

4. grateful

5. appreciate

B) Read the statements and write True or False next to each statement. (5 points)

6. *Both* and *the other* one are expressions we use to make comparisons. _____

7. A fact and an opinion are the same. _____

8. It is a fact that American cars are the best in the world. _____

9. Japan is a beautiful country. This is an opinion. _____

10. The capital of Japan is Tokyo. This is an opinion. _____

C) Match the word with the definition. (5 points)

- | | |
|---------------|-----------------------------|
| 11. mess | a. to deal with a situation |
| 12. dangerous | b. hurt |
| 13. clean up | c. not safe |
| 14. harm | d. to make a place cleaner |
| 15. manage | f. dirty condition, filth |

D) Match the underlined word to the synonym. (5 points)

- | | |
|---|--------------|
| 16. Terry <u>trained</u> hard to win the race. | a. powerful |
| 17. Chris is a <u>smart</u> boy. He learns quickly. | b. practiced |
| 18. She did well on the test. All her answers were <u>right</u> . | c. talented |

19. Tracey is a fast and strong swimmer. d. intelligent
 20. They are very skilled volleyball players. e. correct

E) Answer the questions. Write Yes or No. (5 points)

21. The boys *kicked* the door open. Is kicked a verb? _____
 22. Many of our drinks have *artificial* color. Is artificial an adjective? _____
 23. He was late for the bus so he ran *quickly*. Is quickly an adjective? _____
 24. Sam and I *ride* the bus home from school everyday. Is ride a verb? _____
 25. Reading books is my favorite hobby. Are books and hobby nouns? _____

F) Read the following statements and check ✓ the correct one. (6 points)

26. Did he went ice-skating yesterday?
 Did he go ice-skating yesterday?
 27. My brother plays sumo as a hobby.
 My brother does sumo as a hobby.
 28. Will they go snowboarding this winter vacation?
 Will they play snowboarding this winter vacation?
 29. My sister and I love going swimming.
 My sister and I love doing swimming.
 30. Our family goes skiing every year in Nagano.
 Our family plays skiing every year in Nagano.
 31. Have you ever played surfing?
 Have you ever gone surfing?

G) Read the text. Write the correct letter to complete each sentence. (14 points)

Save the Sea Turtles

Sea turtles live in the ocean. They are endangered. Thousands of turtles die each year. Some of them get caught in fishing nets. Hotels, houses, seawalls and other buildings on the beach are a problem for sea turtles. Sea turtles lay their eggs in the sand on beaches. People and buildings on the beach can scare turtles too. This can stop mother turtles from laying eggs. Sea turtles often hide in seaweed, but **seaweed** also holds pollution such as oil. The oil can make turtles sick. Garbage in the ocean like balloons and plastic bags are also dangerous for turtles. In addition, people kill turtles. They eat turtles and turtle eggs. They also want to use their shells for decorations. This is illegal. People around the world want to **protect** sea turtles. Volunteers visit

beaches to keep them safe for mother turtles and their eggs. Police try to stop people from killing turtles for their meat, eggs and shells. Scientists are studying sea turtles to help us understand them better. In Baja, Mexico, scientist Jay Nichols follows sea turtles with satellites and computers. In 1996, he **released** a sea turtle named Adelita with a radio transmitter on her back. The turtle traveled 6000 miles to Japan! We have a lot to learn about caring for these amazing animals.

[**seaweed** - a plant that grows in the ocean; **protect** - to defend against harm or loss; **released** - set free]

- ___ 32. Sea turtles are ____.
- a. dangerous
 - b. endangered
 - c. danger
- ___ 33. One reason sea turtles die is ____.
- a. they get cold in the water
 - b. they try to swim too far
 - c. they eat balloons and plastic bags
- ___ 34. Sea turtles lay their eggs ____.
- a. in the sea
 - b. in the sand
 - c. in seaweed
- ___ 35. People kill sea turtles because people ____.
- a. want to use their shells for decorations
 - b. think they are dangerous
 - c. want to use their teeth
- ___ 36. People want to protect sea turtles, so they ____.
- a. build hotels, houses, seawalls and other buildings on the beach
 - b. visit beaches to protect mother turtles and their eggs
 - c. eat turtle meat
- ___ 37. Scientists are studying sea turtles because ____.
- a. they want to use satellites and computers
 - b. they wanted to see if a turtle could swim 6000 miles.
 - c. they want people to understand them better
- ___ 38. What is the writer's opinion? Sea turtles ____.

