An integrated approach to establish validity and reliability of reading tests

Salim Razı
Çanakkale Onsekiz Mart University, Turkey

Abstract

This study presents the processes of developing and establishing reliability and validity of a reading test by administering an integrative approach as conventional reliability and validity measures superficially reveals the difficulty of a reading test. In this respect, analysing vocabulary frequency of the test is regarded as a more eligible way of measuring validity. A study was conducted at Dokuz Eylül University and Çanakkale Onsekiz Mart University with three colleagues and 100 undergraduate students to establish validity and reliability along with readability and vocabulary frequency of a 32-item reading test which was developed by the researcher. Such detailed assessment is highly recommended for researchers who are in need of preparing pre and post tests which are different from each other.

Keywords: *assessing reading, reliability, validity, multiple choice, item analysis, item difficulty*

In this article, it might be helpful to identify the differences among the three confusing terminologies of 'assessment', 'evaluation', and 'testing'. As identified by Noda (2003), assessment requires administering examinations to learn about the students' performances along with observing them in the classroom activities; however evaluation has nothing to do with formal examinations since it deals with the students' performances in the classroom during the activities. On the other hand, testing requires administering specifically prepared examinations and is not interested in students' performances in the activities. Fry (1977a) groups comprehension questions in two broad categories as *objective* and *subjective* ones. The former can be regarded as Pearson and Johnson's (1978) *textually explicit* questions; and the latter as *textually implicit* ones. Then, an objective or a textually explicit question provides both information about question and correct answer whereas a subjective or textually implicit question presents the correct answer only through combining a set of related sentences.

It would be wise to remember that it is unfeasible to assess readers' comprehension of the text since reading comprehension "is totally unobservable" therefore requires analyzing 'behaviour' (H. D. Brown, 2001, p. 315). Such analysis depends on several actions such as doing, choosing, transferring, answering, condensing, extending, duplicating, modelling, and conversing. To H. D. Brown, these actions can be observed in acting physically, selecting among options, summarizing the text, responding comprehension questions, outlining, adding

an end to a story, translating into L1, following instructions to assemble a toy, and taking part in a conversation (p. 316).

Alderson (2000) concludes that although reading is regarded as a process, it is quite common to assess readers' comprehension with reference to product not process as product is much easier than process in terms of investigation of readers' comprehension. In this respect, Alderson reveals the most common techniques in testing reading as gap-filling, cloze, multiple-choice, summary, dichotomous-item, editing, question-answer, matching, and ordering tests (See Alderson, 2000 and Razı, 2005 & 2007 for a detailed account of these test types).

Evaluating test quality

To ensure reliability and validity of reading tests which are constructed by the integration of above mentioned techniques, testers refer to a number of various analyses that will be explained below.

Reliability

Noda (2003) indicates reliability as a crucial element of standardized testing and points out that test-takers receive almost the same mark when they are delivered a reliable test for multiple times. This implies that if a reading test is reliable then the tester is sure that the test is consistent and test-takers perform almost the same at all times the test is delivered. Noda highlights that group performance is also another criterion that needs to be taken into consideration while dealing with reliability. If a group of test-takers perform much better or much worse in any test when compared with their previous scores on similar tests, then such a test cannot be regarded as reliable.

The most common ways of assessing reliability is measuring 'stability or test-retest', 'alternate form' (Kaplan & Saccuzzo, 2001), 'internal consistency – Alpha' (Aiken, 2003), and 'interrater reliability or interrater objectivity' (Goodwin, 2001). To measure stability of a test, the tester delivers the same test twice with a probable interval of two weeks and calculates the correlation between these two tests in which reliability is reflected. On the other hand, by producing two versions of the same test in which the items differ from each other very slightly, the tester is able to calculate reliability by working on the correlations between these two tests. Thirdly, internal consistency is also regarded as another crucial element of reliability. Such consistency presumes that a test-taker's performance is similar in items which are similar to each other. Fourthly, interrater reliability reveals the consistency of two or more raters' scores on the same performance.

The marking procedure needs to be quite objective to provide reliability as it is also essential for reliable tests to be marked with almost the same results by different markers (S. Brown, 1994). To provide reliability, test-takers are required to use test techniques which are familiar to the test-takers; otherwise failure may occur as a result of unfamiliarity with the question types which results in an unreliable test. Noda (2003) does not approve administration of a single long lasting test at the end of a course as it decreases reliability of the test; instead she recommends daily evaluations of the readers for reliable results. S. Brown also calls attention to a precarious attempt to increase reliability of tests. She indicates that testers restrict their questions to objectively marked items such as multiple choice tests which in turn results in failure in the test's validity.

Validity

It is possible to regard a test valid if it measures what it is expected to measure in an efficient way (Crocker & Algina, 1986). The most common evidences of validity are 'face', 'content', 'criterion', 'construct' and 'discriminative' and 'generalizability' (Carducci, 2009). *Face validity* compares the test with what it is supposed to be assessing in terms of its facet whereas *content validity* questions the content of the test and compares its appropriateness with the instructional objectives. Moreover, *criterion validity* investigates the scores of the test and compares them to that of an external criterion while *construct validity* aims to match a theoretical concept with the test by following these three steps of specifying theoretical relations, examining empirical relations, and then interpreting them (Carmines & Zeller, 1991). *Discriminate validity* ensures that the test is not related with other instruments excessively (Campbell & Fiske, 1959) and the *validity of generalizability* indicates how appropriate the test is to test-takers in a variety of settings.

Validity is supposed to be more important than reliability as a reliable test may not be valid. For example, a reliable reading test which consists of gap filling questions on grammatical items cannot be regarded valid for assessing reading comprehension. Noda (2003) notes that the texts and the tasks in the test are the factors which identify validity of the test and she considers *independence of modalities* as an important element which implies that testers need to isolate the tested language skill from the others. Unfortunately, a considerable number of reading professionals prefer to integrate the other language skills into reading tests as it is quite common to encounter summary questions followed by a text. In such cases a very crucial question arises: "What is the aim of the tester?" If the answer is testing reading comprehension then is this an effective way of testing readers' comprehension on a productive skill of writing? Therefore, such tests cannot be considered to be valid.

Standard error

Basically readers are categorised as *good* and *poor* ones; moreover it is also possible to add one more group of readers to these two namely *mediocre*. Good readers are expected to achieve higher results whereas poor ones are expected to achieve lower results. However, mediocre readers are expected to survive if they are delivered valid and reliable test. In this respect, standard error identifies their possibility of survival, in other words being successful in the test. Noda (2003) considers administration of a single long lasting test at the end of a course as an ill-inspired attempt as standard error cannot be taken into consideration in such a single-test.

Readability analysis

Readability scores aim to measure the linguistic complexity of texts (Alderson, 2000) and to materialize this a number of readability formulas have been developed to assess the text's difficulty by considering them as products (Wallace, 1992) with reference to the lengths of words and sentences in them (Fry, 1977b). For example, Fry's formulate works on a sample of 100 words which come from the beginning, middle, and the end of the text; and calculates the difficulty in positive correlation with word and sentence lengths. There are also formulas which aim at estimating lexical load by identifying frequencies of words that appear in a text or by examining their lengths. Another approach to assign readability of a text is investigating the sentence lengths in it. However, Alderson regards it as a controversial issue since adding new words to a sentence may simplify its comprehension. Alderson concludes that it is almost impossible to identify the difficulty of a text absolutely, therefore he recommends use of authentic texts in appropriate to the aim.

However, Chastain (1988) revises the validity of readability analysis and reveals that it would be unwise to blame linguistic complexity on its own for reading comprehension problems as the process of reading is regarded as an interactive one in which readers' schemata and their interest in reading the text are considered to be major contributors to the understanding of the texts. Wallace (1992) argues that also reduced clauses need to be regarded since they shorten sentences by creating difficulty. Alderson (2000) also expostulates the use of readability analysis as he regards it as a product approach to reading with the two limitations of variation in the product and also method which is used to measure the product.

Corpus linguistics

Although definition of a corpus regards any collection which includes more than one text, in relation to modern linguistics the four characteristics of 'sampling and representatives', 'finite size', 'machine-readable form' and 'a standard reference' should also be incorporated in corpus studies (McEnery & Wilson, 1996).

Conrad (2005, p. 394) reveals that the corpus is constituted of both written texts and transcriptions of speeches. She calls attention to the importance of authenticity of the materials in the corpus as it is a "collection of naturally occurring texts that is stored in electronic form" rather than the materials which are prepared for teaching language. Conrad maintains that technological advances enabled to achieve large scale corpora consisting of hundreds of millions of words compared to one-million word corpora in the 1970s. Such advances encourage dictionary writers to give frequency of words. It was Frith (1957) who first introduces the term of collocation; however, his proposal is materialized by the advances in corpus linguistics. Such advances undoubtedly assist Lewis (1993) to give birth to *the lexical approach* where the emphasis is on building lexical units. Richards and Rodgers (2001) indicate that apart from collocations, binomials, trinomials, idioms, similes, connectives, and conversational gambits also appear in language.

Bias and testing reading

As discussed earlier, any quality test is required to be valid and reliable along with an acceptable standard error value. Besides, bias can be regarded as a concept to be removed from a quality test (Murphy, 1994) since it prevents testers to evaluate test-takers' responses in a fair way. In order to identify whether the anomalous looking question is biased or not, Murphy recommends testers to examine test-takers' responses by determining any 'atypical' performances. To make the concept more comprehensible, Murphy gives an example from Hannon and McNally (1986) where they examine a biased reading question as presented below.

An example from the reading text:

The man was very late and just managed to jump **. . .** the bus as it was pulling away from the stop.
1 at
2 up
3 on
4 by

(Murphy, 1994, p. 297)

Over half of the test-takers failed to choose the correct answer for the above mentioned question because of their insufficient knowledge of colloquial English, rather than the inability in reading comprehension. An interesting conclusion on biased results comes

from Capel, Leask, and Turner (1995) who indicate that multiple choice questions as in Hannon and McNally' (1986) example, seem to leverage males over females.

The study

Testers generally aim at establishing reliability and validity for their tests by administering the analyses that were discussed above under the subtitles of reliability and validity. However, if the aim is testing reading, then testers also intend to refer to various readability analyses to identify the difficulty of the texts in their tests. Readability formulas have long been criticised since they merely take into consideration word and sentence lengths (Wallace, 1992). Then, apart from readability analyses, there arouses a need to investigate the other aspects in the text. In this respect corpus linguistics studies may assist reading testers.

Although reliability and validity analyses are regarded as standard procedures, calculating word frequency is not taken into consideration. Therefore, the present study aims to establish validity and reliability along with readability and vocabulary frequency of a reading test which was developed by the researcher. In this respect, the researcher aims to produce a more reliable and valid reading test. Therefore, the present study aimed to answer whether it was possible to evaluate reading tests in terms of vocabulary frequency and integrate this with the other means of reliability and validity measures.

The student participants of the study were instructed to answer questions in 90 minutes. They were also reminded that their wrong answers did not have any impact on their score from that test. Besides, they were not allowed to use dictionaries during the test.

Setting

The validity was conducted in the ELT Departments of Dokuz Eylül University and Çanakkale Onsekiz Mart University with three colleagues whereas the reliability was conducted in the ELT Department of Çanakkale Onsekiz Mart University with a number of 100 undergraduate students over the fall semester of the 2008-2009 academic year. ELT Department was suitable for this study because of the high English language proficiency of the participants.

Participants

The study consists of 100 students from preparatory, freshman, sophomore, junior, and senior classes at the average age of 20. All the participants were considered advanced Turkish learners of English as they had to take the placement test of Foreign Language Examination (YDS) which is administered once every year by Higher Education Council Students Selection and Placement Centre of Turkey (ÖSYM), to study at the ELT Department. Apart from YDS, in order to enrol first year courses, the students were required to take an exemption examination on the registration of the department which tested their proficiency in English by dealing with all language skills along with grammar and vocabulary.

As the department of ELT is a female dominant one, a vast majority of the participants were females. Gender distribution of the participants in the study is shown in Table 1.

Table 1
Gender Distribution of Participants

Classes	Female	Male	Class Total
Preparatory	16	4	20
Freshman	14	6	20
Sophomore	15	5	20
Junior	15	5	20
Senior	14	6	20
Total	74	26	100

Moreover, oral permission had previously been sought from the students to use their test results for research purposes. On this occasion, they were reminded that the data to be collected was for research purposes only; it would be kept confidential, and would have no bearing on assessment of their courses.

Instrument

A four-section, 32-item reading test was developed by the researcher to test reading comprehension. There were four-option multiple choice questions in the first, third, and fourth sections of the test. Such questions were a combination of Pearson and Johnson's (1978) textually explicit, textually implicit, and scriptally implicit questions along with Fry's (1977a) objective and subjective questions. The second section of the test presented paragraph matching questions. As proposed by Alderson (2000), there were more options in the matching section than the task demanded. All the texts in the test were taken from real life reading materials and adjusted for the test. All the questions in the test were prepared by the researcher. The reading test was very similar to University of Cambridge Local Examinations Syndicate (UCLES) Examinations in English as a Foreign Language Certificate of Proficiency in English (CPE) Reading Paper, apart from the replacement of a section.

Findings and discussion

Validity of the reading test

To avoid producing test items which do not require reading the text as proposed by Hadley (2003), the multiple-choice questions were answered without reading the test by an Associate Professor at the ELT Department of Dokuz Eylül University. Then, to provide other validity measures of the reading test, the questions and the texts in the reading test were evaluated by the same colleague in terms of their *content*, *face*, and *criterion-related* validities. Since the questions in the test focused on a variety of aspects regarding reading comprehension such as 'implication', 'opinion', 'main idea', 'detail', 'attitude', 'cohesion', 'coherence', 'text structure', 'global meaning', 'comparison', and 'reference' in either multiple-choice or multiple-matching style, the test was regarded to be valid in terms of its content. Moreover, as the participants of the study were familiar with such texts and question types, it was also valid in terms of its face. As the reading test was quite similar to UCLES CPE Reading Paper, apart from the replacement of a section in accordance with the aim of the researcher, it was regarded valid in terms of criterion-related test.

The reading test was also evaluated by two native English speaking colleagues of Çanakkale Onsekiz Mart University, one of whom employed as an Instructor of English at the Department of ELT and the other employed as an English Language Specialist. Both the texts and the questions in the test were proofread and also the texts were ranked from 1 to 10 according to their difficulty. These two native speakers' recommendations on the language of

the texts and questions were taken into consideration. Besides, the mean values of the two native speakers' text difficulty scores gave an overall idea about the difficulty of the texts which are indicated in Table 2.

Table 2
Text Difficulty Evaluation of Native Speakers

Reading Test		Text Difficulty		
		Native Speaker 1	Native Speaker 2	Mean
	Text 1	8	8	8
	Text 2	9	8	8.5
Part 1	*Text 3*	7	5	6
	Text 4	6	5	5.5
	Mean	7.5	6.5	7
Part 2		8	6	7
Part 3		10	8	9
Part 4		7	7	7
Mean		8.13	6.88	7.5

The native speakers' evaluation of the texts indicates that the language of the texts shows a difficulty level ranging from 5 to 10 on a difficulty scale of 10. The two native speakers' evaluation of the texts shows a high and significant correlation ($r = .782$; $p < .05$). Although there are some slight differences between the difficulty levels of the texts in different sections of the test, this does not affect the validity of it since each section functions independently it the test. To conclude, an overall score of 7.5 on a 10 point scale may indicate that the test is appropriate to be used at proficiency level.

Moreover, readability analyses were administered for each text in the reading test by using Microsoft® Word for the scores of counts and averages. Readability analyses were presented with the results of standard tests namely Flesch reading ease and Flesch-Kincaid grade level which were calculated by using Microsoft® Word. Besides Fog scale level was calculated online at http://www.readabilityformulas.com/free-readability-formula-assessment.php along with and SMOG (Simple Measure of Gobbledygook) readability formula which was calculated online at http://www.harrymclaughlin.com/SMOG.htm.

Table 3 presents the readability scores of the texts along with the details on counts and averages and it indicates that the texts in the reading test consist a total of 4,068 words in four parts. Readability analyses were presented with the results of standard tests namely Flesch reading ease, Flesch-Kincaid grade level, Fog scale level, and SMOG readability formula. Firstly, Flesch reading ease scores which measure readability by using the average sentence length and the average number of syllables per word indicate similarities among the texts in the test. As higher rating scores indicate the easiness of texts and the scores between 30 and 49 are considered to be difficult in Flesch reading ease scale (McLaughlin, 1969); all the texts are attributed to be difficult with reference to Flesch reading ease scores. However, Flesch reading ease scores are attributed to be most reliable for upper elementary and secondary reading materials.

Secondly, Flesch- Kincaid grade level indicates the grade level of a text by measuring textual difficulty and the scores above 12 are demonstrated as 12 in Flesch- Kincaid grade level, Table 3 points out that all the texts in the reading test appear at the level of 12 or above. It is worth to mention that Flesch- Kincaid grade level stands for a grade-school level. Therefore, like Flesch reading ease scores, Flesch- Kincaid grade level scores are also considered to be reliable for upper elementary and secondary reading materials.

Table 3
Scores of Readability Analyses

| | Readability Analyses | Reading Test | | | | | | | | |
| | | Part 1 | | | | | | | | |
		Text 1	Text 2	Text 3	Text 4	Part 1	Part 2	Part 3	Part 4	Total/Mean
Counts	Words	247	265	279	215	1006	1109	708	1245	4068
	Characters	1188	1470	1451	1152	5261	5827	3652	6198	20938
	Paragraphs	5	4	4	3	16	18	8	10	52
	Sentences	10	13	10	10	43	54	30	58	185
Averages	Sentences per paragraph	2.5	4.3	3.3	5.0	3.78	3.6	4.2	6.4	4.49
	Words per sentences	24.2	20.2	27.6	20.7	23.18	20.1	23.2	21.4	21.97
	Characters per word	4.7	5.3	5.0	5.1	5.03	5.1	5.0	4.8	4.98
Readability	Passive Sentences	20%	30%	50%	0%	25%	20%	6%	15%	16.5%
	Flesch reading ease	49.0	30.1	38.7	37.4	38.8	36.2	42.4	40.7	39.53
	Flesch-Kincaid grade level	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0
	Fog scale level	14.10	16.94	12.63	9.11	13.2	13.84	15.20	12.41	13.66
	SMOG readability formula	14.49	15.53	14.75	15.85	15.16	15.14	15.77	15.14	15.30

Although the scores of two readability analyses of Flesch reading ease and Flesch-Kincaid grade level provide a general idea about the texts, they cannot be considered appropriate at proficient level. Therefore, subsequent analyses are required such as the third analysis of Fog scale level which is mainly used to measure readability of non-educational texts. Similar to the Flesch scale, the Fog scale also compares syllables and sentence lengths and words with three or more syllables are considered to be 'foggy'. Fog scale level scores indicate that the texts are hard and almost difficult to understand which makes it an appropriate instrument for proficient level of EFL learners.

Moreover, a fourth readability analysis of SMOG readability formula was administered to predict the difficulty level of texts. Like the Fog scale, the SMOG formula also identifies words which have three or more syllables as polysyllabic which make the text difficult to read. The average SMOG level of the texts indicates that, the reading test is at a level between college and university degree with reference to the scale provided by McLaughlin (1969). This score also makes the reading test an appropriate instrument to test reading comprehension at proficient level.

The scores of readability analyses gave a clear picture of the texts' difficulty levels by examining them with reference to linguistic features. However, the nature of such readability analyses does not allow the contextual investigation of lexical items in the text. Unavoidably,

such a factor plays a crucial role in reading comprehension. Therefore, the lexical items in the reading test were also evaluated.

To enable this evaluation, all the vocabulary in the texts of the reading test was listed except for numbers and proper nouns. Repetitive occurrences of existing words were not taken into consideration. Then, these words in the list were ranked according to their frequency of usage by the help of a computer programme WordCount™ which presents the 86,800 most frequently used English words by ranking them in an order of commonness where the data is based on the British National Corpus®. The words which do not appear in WordCount™ were ranked in the 86,801st place in the list. Table 4 presents the mean values of frequency of the words in the reading test.

Table 4
Mean Value of Frequency of Words in the Reading Test

Reading Test		Frequency of Words
	Text 1	3009.24
	Text 2	3438.70
Part 1	*Text 3*	2261.30
	Text 4	2517.53
	Mean	2806.70
Part 2		6740.02
Part 3		3399.97
Part 4		3987.75
Mean		4233.61

Table 4 above reveals that on average the words appear in a frequency rank of 4234 in the reading test. This average score implies that the texts include less frequently used words along with very common ones. Moreover, the frequencies of the words in the test show high and significant correlations between Part 1 and Part 2 ($r = .503$; $p < .01$); Part 1 and Part 3 ($r = .545$; $p < .01$); Part 1 and Part 4 ($r = .840$; $p < .01$); Part 2 and Part 3 ($r = .625$; $p < .01$); Part 2 and Part 4 ($r = .824$; $p < .01$); and Part 3 and Part 4 ($r = .439$; $p < .01$).

Table 5 displays the evaluation scores of the reading test for its validity in terms of difficulty levels of native speakers, readability scores, and word frequency analyses.

Table 5
Reading Test Validity Evaluation

Reading Test		Native speaker			Readability				Word frequency
		1	*2*	*Mean*	*Flesch*	*Flesch-Kincaid*	*Fog*	*SMOG*	
	Text 1	8	8	8	49.0	12.0	14.10	14.49	3009.24
	Text 2	9	8	8.5	30.1	12.0	16.94	15.53	3438.70
Part 1	*Text 3*	7	5	6	38.7	12.0	12.63	14.75	2261.30
	Text 4	6	5	5.5	37.4	12.0	9.11	15.85	2517.53
	Mean	7.5	6.5	7	38.8	12.0	13.20	15.16	2806.70
Part 2		8	6	7	36.2	12.0	13.84	15.14	6740.02
Part 3		10	8	9	42.4	12.0	15.20	15.77	3399.97
Part 4		7	7	7	40.7	12.0	12.41	15.14	3987.75
Mean		8.13	6.88	7.5	39.53	12.0	13.66	15.30	4233.61

To conclude with reference to Table 5, four parts of the reading test show similarities in terms of the scores of difficulty levels of native speakers, readability analyses, and word

frequency levels. The scores indicate it as an appropriate material to be used with proficient readers of EFL; therefore it can be considered to be valid.

Reliability of the reading test:

To test the reliability of the reading test, item analysis was employed to the 32-questioned reading test which was administered to a group of 100 participants in the department of ELT for item analysis in terms of *item difficulty* and *item discrimination*.

To administer item analysis process, first the participants' answers were marked by the researcher. The marking process was completely objective since it was done by computer. To enable this, the researcher formulized an Excel spreadsheet to feed the data into computer. In this respect, the correct answers were given '1' point where the wrong ones were given '0' point. As all the items were totally objective in terms of marking process, there was no need for an interrater reliability score. Then the participants' total scores' were listed in descending order. The answers of the 27 participants who were at the top of the list and the 27 participants who were at the bottom of the list were taken into consideration in the next step. Later each item in the reading test was calculated in terms of correct answers in the top 27-participant group and in the bottom 27-participant group.

To calculate *item difficulty* the number of correct answers in the top 27-participant group was added to the number of correct answers in the bottom 27-participant group. The sum was divided by 54 and indicated the item difficulty score for each item in the reading test.

On the other hand, to calculate *item discrimination*, the number of correct answers in the bottom 27-participant group was subtracted from the number of correct answers in the top 27-participant group. The amount was then divided by 27 and indicated 'item discrimination'. Table 6 shows the rationale used for the evaluation of the items in the reading test.

Table 6
Rationale for the Item Analysis Process

Group	(p) Item Difficulty	(r) Item Discrimination	Interpretation
1	>0.90	No value	Preferable if teaching process is effective
2	0.60-0.90	>0.20	Practically appropriate item
3	0.60-0.90	<0.20	Needs to be revised
4	<0.60	>0.20	A formidable but discriminative item: Appropriate for high standards
5	<0.60	<0.20	A formidable but non-discriminative item: Needs to be removed

The 32 items in the reading test were evaluated with reference to the rationale presented in Table 6. The results in Table 7 indicate that all the items in the reading test, except from the items 25 and 29 were appropriate to be used in the test. Therefore, these two items were removed from the reading test. The answers of the participants on the remaining 30 items were then analyzed to find out the reliability of the reading test. Reliability analysis revealed a Cronbach's alpha score of $\alpha = .81$ over 30 items in the reading test. This score indicates that the 30-question reading test is acceptably reliable. Table 7 presents the results of reading test on item analysis in terms of 'item difficulty' and 'item discrimination'.

Table 7
Item Analysis of the Reading Test

Items	(p) Item Difficulty	(r) Item Discrimination	Group
Item 1	0.796296	0.407407	2
Item 2	0.870370	0.259259	2
Item 3	0.796296	0.407407	2
Item 4	0.851852	0.296296	2
Item 5	0.777778	0.444444	2
Item 6	0.740741	0.444444	2
Item 7	0.611111	0.703704	2
Item 8	0.796296	0.407407	2
Item 9	0.629630	0.592593	2
Item 10	0.648148	0.333333	2
Item 11	0.611111	0.629630	2
Item 12	0.611111	0.259259	2
Item 13	0.611111	0.407407	2
Item 14	0.611111	0.333333	2
Item 15	0.814815	0.370370	2
Item 16	0.648148	0.555556	2
Item 17	0.777778	0.296296	2
Item 18	0.611111	0.481481	2
Item 19	0.759259	0.407407	2
Item 20	0.722222	0.333333	2
Item 21	0.629630	0.592593	2
Item 22	0.611111	0.259259	2
Item 23	0.685185	0.629630	2
Item 24	0.722222	0.555556	2
Item 25	1	0	1
Item 26	0.611111	0.481481	2
Item 27	0.740741	0.296296	2
Item 28	0.759259	0.259259	2
Item 29	0.462963	0.111111	5
Item 30	0.740741	0.444444	2
Item 31	0.740741	0.518519	2
Item 32	0.648148	0.555556	2

Conclusion

This paper includes information about establishing the reliability and validity of a reading test, as well as a description of the development procedure of the test. After such detailed validity and reliability analyses, it might be possible to report about a reading test's restrictions, such as readability of the texts, what grades the test is appropriate for, and the how discriminative the questions in the test are.

The study aimed at describing the process of establishing validity and reliability of a reading test in detail with the intention of providing valuable information about multiple assessment criteria both to teachers of reading who rely on reading tests to determine reading skills of their students and researchers who are in need of reliable reading assessment tools for their pre and post tests. Establishing such validity and reliability analyses might also be beneficial for testers as they depend on assessment tools for making decisions about the candidates.

In order to offer any opinions about the quality of a reading test, some assessment criteria are supposed to be administered. Assessing any reading test with just a single criterion may not hinder realistic results. Therefore, evaluating reading tests in terms of multiple factors may assist teachers, researchers, and testers to decide for themselves which reading test is most appropriate for their particular needs.

The general tendency to assess a reading test is dealing with its validity and also reliability. Such an assessment requires reading tests which are free of bias and distortion. However, such analyses do not necessarily reveal exact difficulty of the texts in the test as reliability focuses on question items rather than the texts in the test. In addition to these two, calculating readability also gives an idea about the difficulty of a text. Nevertheless, readability analyses can also be considered superficial as they merely deal with either word or sentence lengths. Then, there arouses the necessity of scrutinizing the words in the texts of a reading test. Therefore, vocabulary frequency analysis may assist testers to assess their texts more deeply.

Implications

Such detailed assessment of a reading test in terms of its validity and reliability is highly recommended for researchers who are in need of preparing pre and post tests for experimental studies. Then, they will be able to administer pre and post tests which are both different from and identical to each other. However, it might be very tiring for reading teachers to administer such detailed analysis for their reading tests.

Due to their profession, researchers might be aware of the importance of establishing validity and reliability for their reading tests; however, this may not be the case for teachers as their principal goal is teaching rather than researching. Nevertheless, teachers should also be encouraged to use valid and reliable tests to assess their students' reading skills. It might be beneficial to assist reading teachers at any grade to achieve this goal by the help of in-service training.

In case of failure in providing in-service training to professionals on assessing validity and reliability of reading tests, it might be beneficial to form databases which constitute of valid and reliable reading tests. Being able to have an access to such databases will allow teachers, researchers, and also testers to select the most appropriate reading test in accordance with their needs. As cooperation with colleagues is one of the essential elements of establishing validity of a reading test, such collaboration among colleagues should be encouraged to establish more valid reading tests.

Doubtless, the process of identifying vocabulary frequency in a reading test is both tedious and time-consuming. Therefore, computer programmers can be encouraged to add a feature to their word processors to calculate vocabulary frequency of a reading test which is very similar in principle to calculating reliability of a text in Microsoft Word®. Then, the easiness of receiving vocabulary frequency level may also encourage reading teachers to assess their texts also in terms of vocabulary frequency.

Moreover, in order to evaluate frequency of vocabulary scores consistently, there is a need of developing sample criteria. Then, further researchers may calculate vocabulary frequency of a variety of texts from a broad range, and correlate them with different levels of language learning.

Biographical information

Assist. Prof. Dr. Salim Razı works at the ELT Department of Canakkale Onsekiz Mart University, Turkey. His research interests include schema theory, reading comprehension, assessing reading, metacognitive reading strategies, academic writing, and research culture. He can be contacted at salimrazi@gmail.com.

References

Aiken, L. R. (2003). *Psychological testing and assessment* (11th ed.). Boston: Allyn and Bacon.

Alderson, J. C. (2000). *Assessing reading.* Cambridge: Cambridge University Press.

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy.* New York: Longman.

Brown, S. (1994). Assessment: A changing practice. In B. Moon, & A. Shelton Mayes (Eds.), *Teaching and learning in the secondary school* (pp. 266-272). London: Routledge.

Campell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Capel, S., Leask, M., & Turner, T. (1995). *Learning to teach in the secondary school: A companion to school experience.* London: Routledge.

Carducci, B. J. (2009). *The psychology of personality* (2nd ed.). Chichester: Wiley-Blackwell.

Carmines, E. G., & Zeller, R.A. (1991). *Reliability and validity assessment.* Newbury Park: Sage Publications.

Chastain, K. (1988). *Developing second-language skills theory and practice* (3rd ed.). Orlando: Harcourt Brace Jovanovich.

Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393-409). New Jersey: Lawrence Erlbaum.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Forth Worth: Harcourt Brace Jovanovich.

Frith, U. (1985). Beneath the surface of developmental dyslexia. K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Neuropsychological and cognitive studies of phonological reading* (pp. 301-303). London: Lawrence Erlbaum Associates.

Fry, E. B. (1977a). *Elementary reading instruction.* New York: McGraw-Hill.

Fry, E. B. (1977b). Fry's readability graph: Clarification, validity, and extension to level 17. *Journal of Reading, 20,* 242-252.

Goodwin, L. D. (2001). Interrater agreement and reliability. *Measurement in Psychical Education and Exercises Science, 5,* 13-14.

Hadley, A. O. (2003). *Teaching language in context* (3rd ed.). Boston, Massachusetts: Heinle & Heinle Publishers.

Hannon, P., & McNally, J. (1986). Children's understanding and cultural factors in reading test performance. *Educational Review, 38,* 237-246.

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principle, applications and issues* (5th ed.). Belmont, CA: Wadsworth

Lewis, M. (1993). *The lexical approach.* Hove: Language Teaching Publications.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics.* Edinburgh: Edinburgh University Press.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading, 12,* 639-646.

Murphy, P. (1994). Assessment and gender. In B. Moon & A. Shelton Mayes (Eds.), *Teaching and learning in the secondary school* (pp. 287-298). London: Routledge.

Noda, M. (2003). Evaluation in reading. In H. Nara & M. Noda (Eds.), *Acts of reading: Exploring connections in pedagogy of Japanese* (pp. 197-222). Honolulu: University of Hawai'i Press.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension.* New York: Holt, Rinehart and Winston.

Razı, S. (2005). A fresh look at the evaluation of ordering tasks in reading comprehension: Weighted marking protocol. *The Reading Matrix, 5,* 1-15.

Razı, S. (2007). *Okuma becerisi öğretimi ve değerlendirilmesi* [Teaching and assessing reading]. İstanbul: Kriter Yayınları.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching.* Cambridge: Cambridge University Press.

Wallace, C. (1992). *Reading.* Oxford: Oxford University Press.

Appendix:

Part 1

You are going to read four extracts which are concerned in some way with drug addiction and TV programmes. For questions 1-8, choose the answer (A, B, C or D) which you think fits best according to the text. (24 marks)

Drug addiction and drug abuse

There is a subtle but important difference between drug abuse and addiction. Someone can abuse drugs without being addicted but the opposite is not true. Some experts have outlined different stages in the addiction process.

Stage one is the exploratory phase, stage two is the recreational, stage three is the abusive, and stage four is the dependent that is associated with the word addiction.

Drug addiction implies a loss of choice where the need is paramount. Drug abuse implies the implementation of a choice to abuse a substance. This choice is born out of a desire to use the substance to help ease circumstances or situation. It is one step further on the road to addiction because it is no longer being used for purely recreational reasons.

It is important to pay attention to drug abuse because it has within it the seeds of the compulsivity that leads to dependency and addiction. This is true in most cases though not in all. There are people who can abuse drugs but are impervious to becoming dependent.

Drug addiction means that drug abuse has become a compulsive need for which there is only one solution. At this point the person feels that there is no choice and that the urge is beyond their control. There are two kinds of addictions: physical and psychological. Often they occur simultaneously because the body is not naturally divided at mind and body.

1 It can be inferred from the text that

A it is possible to be addicted to drugs without abusing them.

B professionals do not deal with the distinction between drug abuse and drug addiction.

C drug addiction is the final step which is preceded by drug abuse.

D a drug dependent cannot be regarded as addicted to drugs.

2 The writer of the text concludes that

A every drug abuser results in addiction.

B a drug addict can control the feelings not to take drugs.

C psychological addiction precedes physical addiction.

D a psychological addict is also a physically addict.

Adapted version of this original text retrieved from Gandino State Medical University,
www.gramb.by/file/kafedry/socov/uch_gas.doc

READING

Test

Full Name   :

Number   :

Class   :

INSTRUCTIONS TO CANDIDATES

Do not open this booklet until you are told to do so.

Write your name, number, and class on the answer sheet in the spaces provided.

There are thirty-two questions in this paper.

Answer all questions.

Mark your answers on the answer sheet.

You may write on this question paper, but you must transfer your answers to the answer sheet within the time limit.

INFORMATION FOR CANDIDATES

You have 90 minutes to finish the test.

Wrong answers will not affect your score.

Answers which do not appear on the answer sheet will not be taken into consideration.

This question paper consists of 12 printed pages.

Media and Children

Families can use a number of tools to be selective about the kinds of TV programmes or movies they watch. There are some good websites that offer reviews or ratings about the content of movies, videos, and DVDs. Although the existing industry ratings can be a useful general guide for parents, the sites 'Common Sense Media', 'MediaWise Kidscore', and 'Kids in Mind' offer more detailed descriptions of content that can help parents select or discuss movies with their children. All of them are free, and information about the sponsoring organization and the methods they use to rate content are clearly described.

Why would parents use the websites presented above when the television industry already provides ratings for TV shows? These ratings could be used to help guide choices, and to programme V-chips that are in all recently made TV sets, but this assumes that the ratings are valid and reliable. One study found that a panel of parents, grandparents, and professionals often agreed with industry ratings of whether content was *inappropriate* for children (e.g., all "R-rated" movies and TV-MA rate television programmes), but they often disagreed on whether movies or programmes rated as being appropriate for children really were.

A more recent national survey found that only half of all parents think that most TV shows are accurately rated. Most of the parents who do use ratings found them at least somewhat useful, however, many parents do not know what the ratings mean. Given this discrepancy, the more detailed information provided by the websites described above allows parents to make their own judgements about whether a television programme or film is appropriate for their own children.

5   The list offered by the writer is different from the others

   A   since it gives a lengthy description of programmes.

   B   since it is free.

   C   because of the methods they use.

   D   as their ratings are accurate.

6   It can be concluded from the text that

   A   parents, grandparents, and professionals always agree with each other on industry ratings.

   B   half of the parents appear to believe in the accuracy of TV shows' ratings.

   C   parents in general are interested in ratings.

   D   TV ratings on their own are sufficient for parents to judge whether a programmes is suitable or not.

Co-occurring disorders

Dual diagnosis, or co-occurring disorders, refer to drug addiction which is accompanied by an emotional or psychiatric illness. Either type of disorder is complex on its own; together, a dual diagnosis will affect the individual socially, spiritually, physically, and psychologically. The interaction of the different components of dual diagnosis can interact so that diagnosis, treatment and recovery are made more difficult.

In addition, accurately assessing the extent of emotional or psychological illness while drug or alcohol addiction are present can become very difficult for the healthcare provider, thereby making an effective treatment plan more difficult to compile. This does not mean however, that treatment for dual disorders is ineffective or unheard of. When treating such a disorder, it is necessary to focus on both issues. Only trying to fix one problem will most likely not result in improvement of either arena.

Perhaps one of the best forms of treatment for co-occurring disorders is what is known as integrated treatment where the patient receives treatment for both mental illness and substance abuse from the same clinician or from a team of clinicians. Basically, the team works together to make sure that the different interventions are brought together. This way the client will see no division between mental health and substance abuse treatment. This eliminates the confusion that can often occur when obtaining treatment in two different centres. In integrated treatment each client has his/her specific program. This way, the individual can move at his/her own pace, thereby resulting in a more effective treatment situation which will hopefully lead to long lasting recovery.

3   The writer indicates that a dual diagnosis

   A   is different from co-occurring disorders in terms of emotion.

   B   has an impact on the addicted person in a variety of ways.

   C   cannot be treated.

   D   may occurs only in case of drug addiction not alcohol addiction.

4   It can be concluded that

   A   the best way for the treatment of a dual diagnosis might be the treatment of the mental illness first.

   B   an integrated treatment cannot be conducted with a single clinician.

   C   in an integrated treatment the client's cooperation with the other patients is essential.

   D   an integrated treatment is based on the principle that there exists no distinction between mental health and substance abuse treatment.

Part 2

You are going to read a short story. Seven paragraphs have been removed from the story. Choose from the paragraphs **A-H** the one which fits each gap **(9-15)**. There is one extra paragraph which you do not need to use. **(28 marks)**

Costing an arm and a leg

Whole, a riveting new documentary by Melody Gilbert is about an increasingly visible group of people who call themselves "amputee wannabes". Wannabes desperately wish to have their healthy limbs removed, and some have succeeded in having it done.

9

My interest in amputee wannabes began several years ago. I was trying to understand why so many people have begun to use the tools of medicine for purposes other than curing illness. I noticed that in the same way that some people said they only felt like themselves after, say, getting sex-reassignment surgery, or even taking Prozac, many wannabes said they would not feel like themselves without an amputation.

10

Gilbert's sensitive film allows wannabes to speak for themselves. Many are so articulate and likable that no matter how difficult you find it to understand their desire, you will come away from the film with sympathy for their strange predicament. Yet perhaps the most disturbing figures in *Whole* are the clinicians. Even as the wannabes admit how baffling they find their own desires, the mental health professionals in the film speak with absolute confidence.

11

This claim is not so much false as incomplete. No formal research studies on treatments for wannabes have ever been undertaken. In fact, nobody really knows much about this condition. Only a handful of articles about it have been published, most of them small case studies in obscure medical journals.

12

Dissenting voices of any kind are largely absent from *Whole*. In her eagerness to document the extraordinary stories her subjects tell, Gilbert has produced a film that uncritically accepts those stories at face value. The patients explain

what this condition is and how it should be treated, and the clinicians obediently nod their agreement. The only sceptical voice in the film comes from Jenny. When Jenny decides she cannot stay married to a man who wants to cut his own leg off, her husband accuses her of being narrow-minded.

13

When I first wrote about this condition in the *Atlantic*, I worried that more people might start to identify themselves as wannabes and seek out amputation. Anyone with a rudimentary familiarity with the history of psychiatry cannot help but be struck by the way that mental disorders come and go.

14

First, the conditions are usually backed by a group of medical or psychological defenders whose careers or reputations depend on the existence of the disorder and who insist that the condition is real. Second, there is usually no hard data about the causes or the mechanism of the condition. Third, no independent lab tests or imaging devices are available to provide objective confirmation of the diagnosis, which is usually made solely on the basis of the narratives and behaviour of their patients.

15

By all indications, the number of people identifying themselves as wannabes is growing. Robert Smith, the Scottish surgeon, has six more acceptable candidates for amputation. A popular wannabe listserv, whose membership was 1,400 two and a half years ago, has 3,670 subscribers today. A group of clinicians at Columbia University has set up a Web site to provide information about the condition. They are redefining it as "Body Integrity Identity Disorder." In the meantime, psychiatrists are no closer to understanding the condition, and they are proposing no therapy other than amputation.

Watching full-length TV programmes on Internet increasingly popular

Watching a favourite show you missed on television on the Internet is increasingly popular, two recent studies show. Horowitz Associates found that 16 percent of high-speed Internet users watched at least one full-length TV program online during a week, double the number from last year. Horowitz just released its report: Broadband Content and Services 2007. The Nielsen Company found that 25 percent of the 1,599 Americans surveyed in October have watched full episodes of a TV program in the past three months. Both studies point to the increasing popularity of full-length streaming video on the Internet.

The Nielsen study notes that ABC.com (50 percent), NBC.com (41), CBS.com (37), and Fox.com (24) were the most watched, with other Internet-based alternatives YouTube (17) and iTunes (15) used less often. Horowitz notes that television is still the preferred delivery platform, with 70 percent of Internet users saying they watch TV online because they missed an episode on television. Some watch a show on TV, and then watched it again on the Internet. Others watched a program because someone recommended it. As NewTeeVee wrote today, the increase in watching full TV shows online bodes well for the networks' streaming strategies and for Hulu, a joint venture between NBC and News Corp.

7 It can be inferred from the text that

A there is an increase in the number of studies investigating the popularity of watching TV programmes on the Internet.

B only high-speed Internet users watch TV programmes on the Internet.

C the main reason for watching a TV programme on the Internet is that you can watch it whenever you want.

D the number of Internet users watching TV programmes on the Internet quadrupled last year's number.

8 The number of people watching TV programmes on the Internet will

A increase but they will also keep their habit of watching TV.

B almost be the same in the following years.

C be so high that it will eradicate TV viewers in a couple of years.

D drop in the future.

Part 3

You are going to read a magazine article. For questions 16-22, choose the answer (A, B, C or D) which you think fits best according to the text. **(28 marks)**

Test anxiety: What it is and how to cope with it

You walk into the exam room...confident that you know the material and can pull off a good grade. You're feeling a little nervous, but not any more than at other times in the past. The test arrives, your hand is a little shaky while you're writing your name down on the answer sheet. The first two questions go fine. Then you read the third question. It seems to be coming at you from about 45 degrees off from what you were expecting... Then it happens... Everything goes blank and even the easy questions you know... You suddenly can't understand let alone answer... Ten minutes before the test is about to end, you start to comprehend some of the questions. You answer some of the easy ones. Even the difficult ones suddenly start to make sense. But it doesn't matter anymore. Time's up...

Exam anxiety is a fairly common phenomenon that involves feelings of tension or uneasiness that occur before, during, or after an exam. Many people experience feelings of anxiety around exams and find it helpful in some ways, as it can be motivating and create the pressure that is needed to stay focused on studying. However, in some cases, anxiety can become so intense that it leads to disruptive symptoms that ultimately lead to a negative impact on one's performance. In these cases, it is important for students to attend to their symptoms and find a way to cope effectively, so that their schooling does not suffer any further.

As a first step, it is important to determine whether the anxiety is "true" test anxiety, or is due to a lack of adequate preparation. The student will need to ensure that he/she spends enough time studying, has adequate study strategies, attends class regularly, and understands the class material. If these issues have been addressed and he/she still continues to experience intense symptoms of anxiety, then it is likely that he/she is suffering from true (or classic) test anxiety, and will need to target his/her particular symptoms directly.

Although anxiety can affect each person in different ways, there are several symptoms that are quite common. Some of these are emotional, which include feelings of fear, disappointment, anger, depression, or helplessness. Other symptoms are more behavioural, ranging from fidgeting or pacing to substance abuse or other self-destructive behaviours. There are also physiological symptoms, which include fast heartbeat, feelings of nausea, headaches, lightheadedness, sweating, and other disruptions in bodily functions. Finally, many people experience cognitive symptoms, such as negative thinking about oneself, racing thoughts, loss of memory, and "blanking" out.

Some of the strategies for coping with exam anxiety are quite practical and relatively easy to implement, such as avoiding caffeine, arriving early to the exam, avoiding people who speak negatively, meeting with the professor to discuss class material, getting a good night's sleep, and reading exam directions carefully. Students will also need to ensure that they are practicing good time management skills and managing their stress on a daily basis through exercise, good nutrition, social support, enjoyable activities, and balance in their lives.

One of the most important components in dealing with exam anxiety is stopping a negative spiral from occurring, which can happen when one sign of anxiety (e.g., trembling hands, negative thoughts about one's performance) leads to a "chain of negative thoughts and images...each feeding on the one before and giving rise to another..." This can lead to an increase in one's anxiety level to the point where he/she can no longer perform at an acceptable level. There are many strategies that can be used to interrupt this cycle, such as breathing deeply; relaxing tense muscles; repeating positive, reassuring statements to oneself; taking a short break from the exam situation; and visualizing oneself doing well.

Exam anxiety can be treated very effectively by continually practicing the above strategies. As some of these may be difficult to learn on one's own, Student Counselling Services provide individual counselling, as well as Exam Anxiety and Relaxation workshops, to aid in this process. For more information, please call Student Counselling Services or visit our office in the Student Union's Building.

A Conditions like social anxiety disorder, post-traumatic stress disorder, attention deficit-hyperactivity disorder, gender identity disorder, multiple personality disorder, anorexia, and chronic fatigue syndrome were once seen as rare or nonexistent, then suddenly they ballooned in popularity. This is not simply because people decided to "come out" rather than suffer alone. It is because all mental disorders, even those with biological roots, have a social component. While these new conditions are very different from one another, they share several important features.

B Finally, there is often a treatment for the condition even in the absence of knowledge about its causes and mechanism. The diagnosis of social anxiety disorder, for example, was driven by the development of profitable medications to treat it, such as antidepressant drugs.

C Kevin, a university lecturer and one of several wannabes featured in the film, had his leg amputated by Robert Smith, a surgeon in Scotland who has amputated the legs of two otherwise healthy people. George Boyer shot his own leg off with a shotgun. Others have used chain saws and homemade guillotines. Why? Nobody really knows, including the wannabes themselves, who often say they have had the desire since they were children. "It's obviously peculiar," admits Kevin. "But knowing it is peculiar and saying it is weird does not do away with the problem".

D You might think that clinicians would want to be certain that all options had been exhausted before recommending that patients have their arms or legs amputated, yet the clinicians in the film do not mention alternative treatments. The only person who expresses a hint of uncertainty is Robert Smith who wonders how the amputations he has performed will be perceived in 20 years.

E I published an article about wannabes for the *Atlantic Monthly* and another on the legality of such amputations with my colleague Josephine Johnston for the academic journal *Clinical Medicine*. It was after reading about wannabes in the *Atlantic Monthly* that Gilbert decided to make her film.

F As clinicians start to diagnose the conditions themselves disorder; the become part of popular discourse. Patients reinterpret their own psychological histories, and their behavior changes to match what is expected of people with the condition they believe they have. "I want you to accept that this condition exists," Baz says emphatically in the film, "and that the *only* way it can be sorted out is psychological treatment".

G Oddly, the film also glides past the sexual aspect of the condition and views it as a problem of identity, like gender identity disorder. In the few medical articles where the condition has been discussed, it is known as "apotemnophilia," because clinicians view it as a paraphilia—a displaced sexual desire like transvestism, voyeurism, and pedophilia. This is because many wannabes are attracted to the idea of themselves as amputees, and some are attracted to other amputees.

H The film features a social worker and clinical psychologist who have counselled Boyer in Florida, as well as Michael First, an academic psychiatrist at Columbia University, who has organized several meetings of wannabes and clinicians. First says that the purpose of these meetings is to "facilitate treatment" for the condition, by which he says he means surgical treatment. His apparent certainty that nothing short of amputation can help these people is underscored by ominous music and a screen shot that reads, "There are no medications or therapies known to help wannabes".

Part 4

You are going to read a magazine article. For questions 23-30, name the places by referring to the text. One place may be used more than once. **(20 marks)**

A glorious experience

The summer of 2000 will forever be for me a season to cherish and a time to remember. It was a glorious and spellbinding period. Beginning May 22, 2000 I travelled to ten countries, including Germany, Namibia, Zimbabwe, Zambia, the Netherlands, Australia, Trinidad, Guyana, Curacao (Netherlands Antilles), Barbados, and Costa Rica, not returning to the United States until the last week of August. I lectured in eight of these countries (nine if you count Curacao) and learned a great deal in all of them. It was a whirlwind of experiences, many of which I am only just now beginning to digest. In this current essay I will provide some background, first hand observations, and insight concerning my travel experiences in Zimbabwe. In fact, of all of my summer travels, only Australia, a country to which I actually led a tour group, surpassed Zimbabwe in terms of length of stay and depth of experience.

With the completion of my Africa Day lecture series in Namibia on May 28, 2000, I caught an Air Namibia flight from Windhoek, Namibia to Victoria Falls, Zimbabwe. After a journey of a little less than two hours, my mission was accomplished. I quickly secured my visa, and stood for the first time on Zimbabwean soil. It was wintertime in Zimbabwe, and the weather was dry and cool. The country was beautiful, the people seemed friendly, and I had the sense of great personal satisfaction that I had realized another dream of a lifetime.

Like Namibia, but even more so, I had wanted to go to Zimbabwe from way back. In fact, after Egypt, Zimbabwe was my favoured African travel destination. Indeed, the ruins of its stupendous stone cities built by the Shona people of northeast Zimbabwe had intrigued me for a long time. In addition to the historical, archaeological, and political aspects of the trip, however, and on a more personal note, my first name, Rumoko, given to me as a university student a long time ago, is in fact a Zimbabwean name.

Zimbabwe, in southeast Africa, is a country of more than eleven million people. More than 95% of its citizens are Black. Most of them, more than seventy percent, are Shona, followed numerically by the Ndebele. Whites and Asians constitute less than five percent of the total population. English is the official language followed by Shona and Sindebele. Most of the Whites are of English origin with more than half of them coming to the country after 1945. There are probably less than 100,000 White people, total, in Zimbabwe today. The country of Zimbabwe finally achieved its independence from White minority rule in 1980.

Geographically, Zimbabwe is bordered by South Africa to the south, Botswana to the west, Mozambique to the east, and Zambia to the north. The capital of Zimbabwe is Harare, in the northeast, a city of more than a million people. The second largest city is Bulawayo with a population of about 700,000 people, mostly Ndebele. Most of my time in Zimbabwe was spent in and around Bulawayo.

The School of African Awareness was the principal sponsor and coordinator of my trip to Zimbabwe. As such, the SAA organized my housing, transportation, lecture schedule, and overall itinerary. The essential goal of the SAA, a non-governmental and non-profit organization launched in Bulawayo, Zimbabwe on Africa Day, May 25, 1997, is to "address issues pertaining to African cultural awareness and self-help and self-reliance. Its main focus is to disseminate information to all those committed to the well being of Africa and its people."

My lectures in Zimbabwe began less than twenty-four hours after my arrival in the country. After securing a taxi and being driven for several hours from Victoria Falls to Bulawayo (where I consumed a hot meal, and caught a night's rest), I spoke the following day at the United College of Teachers. Here, on this day, on which I gave the first of several talks at the college, I spoke to a single class of prospective teachers. Interestingly enough, the college did not even have a history component, and the only reason the lecture materialized at all was through the tireless efforts of Mr. Sibanda. Both the students and the teacher were very receptive, however, and I did a broad-ranging slide-presentation that focused on the African presence globally, ancient and modern. I was to repeat the presentation with minor variations with great success during the course of my stay in Zimbabwe. I tried to inspire the students with the history of African people, and make them proud of themselves. A key component to the success of each presentation was the period allotted to questions and answers that followed every talk. It was a real struggle though, for I was fighting what I perceived to be the strong belief that to embrace Africa was to embrace backwardness, while to embrace Europe was to embrace modernity. Almost all of the students wore western style clothes, consisting of shirts and ties for the men, skirts and nylon stockings for the women. A good deal of the women students wore their hair straightened. These were some of the not so pleasant realities of the trip. I suppose that I, like others, have a kind of idealized vision of what Africa and Africans should be, and it is admittedly disappointing when the vision does not materialize. However, there were those Africans, in the minority just like me, who were, in fact, struggling to realize that vision, and identifying and building with this minority made all of the hard work worthwhile.

In addition to the talks that I gave, I toured the city of Bulawayo extensively, visiting both its townships and

16 What does the writer imply in the introduction paragraph?
A One can be unsuccessful due to insufficient study.
B Difficult questions can never be answered due to insufficient time allocated.
C When you are confused it is almost impossible to continue the exam.
D Failure can be triggered by a difficult question.

17 What does the writer say about exam anxiety in the second paragraph?
A Pre exam anxiety is more common than post exam anxiety.
B The merits of anxiety outweigh the defects.
C Exam anxiety should not be dealt with seriously to get rid of it.
D Successful students do not feel exam anxiety.

18 What does the writer imply about the roots of anxiety?
A Anxiety may not be related with insufficient preparation.
B Anxiety is mainly related with study strategies.
C There is a correlation between anxiety and class attendance.
D A true exam anxiety sufferer has trouble in understanding the class material.

19 It is clear from the text that
A common symptoms of anxiety do not occur together in one person.
B cognitive symptoms exist in case of insufficient preparation.
C emotional symptoms are the rarest ones.
D depression may be an indicator of anxiety.

20 What can't be said about the strategies for coping with exam anxiety?
A In order to work, strategies need to be practiced
B They are transmitted by interaction with other people.
C Avoiding caffeine does not help develop appropriate strategies.
D Interacting with other people increases anxiety.

21 What does the writer say in the penultimate paragraph about negative spirals?
A Trembling hands may result in failure in the exam.
B The level of anxiety is stable in a negative spiral.
C A negative spiral always results in failure in the exam.
D Breathing deeply prevents a negative spiral occurring.

22 The writer concludes that
A it is not essential to consult counselling services to treat anxiety.
B taking drugs is superior to getting professional help.
C none of the methods is completely successful.
D addicted people cannot be prevented suffering from exam anxiety.

Full Name :

Number :

Class :

PART 1 (24 marks)

	1				2				3				4			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
	5				6				7				8			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D

PART 2 (28 marks)

	9				10				11				12				13				14				15			

PART 3 (28 marks)

	16				17				18				19				20				21				22			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D

PART 4 (20 marks)

	23				24				25				26			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
	27				28				29				30			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D

---

its most plush neighborhoods. With the various talks, private meetings, public discussions, TV, radio, and newspaper interviews, every day was a busy one, and I remained fully occupied throughout the course of the trip. Among the most important of the sessions in which I participated were full meetings with the Bulawayo Affirmative Action Committee and the Informal Traders Association. Through these sessions, I was able to gain some kind of understanding concerning the local and national political scenes, and gather some insight into Zimbabwe's economic life. I was also fortunate enough to visit one of the white-owned farms being occupied by the war veterans. These Africans, veterans of Zimbabwe's independence struggle against colonial rule, seemed resolute about holding onto the lands that they are currently occupying. Although they were sorely disappointed when I told them about the manner in which the western media was portraying their actions, their morale was high, and got even higher when I told them of the overwhelming moral support that they enjoyed from African-Americans in general.

One of the great highlights of the entire Zimbabwe trip came on a day that I didn't lecture and was driven far from the confines of Bulawayo. In an emotional ceremony held within the centrality of several villages, attended by the local elders and community residents, and augmented by dancers and drummers, I was warmly received, and officially acknowledged as an African finally returned home. I was presented with a magnificent wooden staff and told that I had finally found my family. It was a wonderful episode, and an experience never to be forgotten. I was so moved emotionally, that when asked to speak at the ceremony, I respectfully, but firmly, declined, as I knew that I would have broken down, and wept like a child.

23 The country to the east of Zimbabwe is
  A Mozambique   C South Africa
  B Zambia   D Bostwana

24 The largest city in Zimbabwe is
  A Victoria Falls   C Harare
  B Bulawayo   D Windhoek

25 The latest city that the writer was in before visiting Zimbabwe is
  A Bulawayo   C Victoria Falls
  B Harare   D Windhoek

26 The first city that the writer had been in Zimbabwe is
  A Harare   C Windhoek
  B Victoria Falls   D Bulawayo

27 The city where the writer first delivered a talk in Zimbabwe is
  A Bulawayo   C Victoria Falls
  B Harare   D Windhoek

28 The city where The School of African Awareness started its facilities is
  A Victoria Falls   C Harare
  B Windhoek   D Bulawayo

29 The writer's native country is
  A Egypt   C the United States
  B the Netherlands   D Zimbabwe

30 The country in which the writer stayed the least on a summer travel is
  A Australia   C Egypt
  B the Netherlands   D Zimbabwe