

Measuring the Measurement: A Preliminary Analysis of the Reading Skills Test Items of College English

Yonas Adaye*

Introduction

Measuring language knowledge is the main concern of language testing. Measuring the language test, designed to assess the reading skills of first year undergraduate students at Addis Ababa University, is the major objective of this study.

Since early 1990s the Department of Foreign Languages and Literature has made a significant change in the teaching of the English language from traditional or teacher-centred to communicative or learner-centred approach. Nevertheless, the content, items, and format of the reading skills test of the course are subject to critical investigation. To date, no research has been undertaken to critically assess the capability of the reading skills test per se to measure the reading skills of the examinees. In this paper, I have attempted to:

- analyse the sample reading skills test items of College English using ITEMAN to find out whether they discriminate among high and low achievers;
- analyse questionnaires filled by instructors of the course to determine the types of constructs, clarity of instructions, relevance of format, content of the passage, and overall assessment of the reading skills test; and
- point out the implications of the study for teaching and testing reading skills in a foreign language context.

* Lecturer, Institute of Language Studies, A.A.U.

Measuring the Measurement: A Preliminary Analysis of the Reading Skills Test Items of College English

Yonas Adaye*

Introduction

Measuring language knowledge is the main concern of language testing. Measuring the language test, designed to assess the reading skills of first year undergraduate students at Addis Ababa University, is the major objective of this study.

Since early 1990s the Department of Foreign Languages and Literature has made a significant change in the teaching of the English language from traditional or teacher-centred to communicative or learner-centred approach. Nevertheless, the content, items, and format of the reading skills test of the course are subject to critical investigation. To date, no research has been undertaken to critically assess the capability of the reading skills test per se to measure the reading skills of the examinees. In this paper, I have attempted to:

- analyse the sample reading skills test items of College English using ITEMAN to find out whether they discriminate among high and low achievers;
- analyse questionnaires filled by instructors of the course to determine the types of constructs, clarity of instructions, relevance of format, content of the passage, and overall assessment of the reading skills test; and
- point out the implications of the study for teaching and testing reading skills in a foreign language context.

* Lecturer, Institute of Language Studies, A.A.U.

The research questions raised Were:

- Do the items of reading skills test of College English measure the reading abilities of the first year students?
- How do instructors perceive the content, format, and specific items of the reading skills test of College English?

The terms below are used with the following tentative definitions in this context:

Item:	an item is a test question or test task.
Item difficulty or (Facility value, FV):	an item difficulty or facility value is measured by the proportion of candidates getting the answer correct compared with those getting it wrong.
Item discrimination or (Discrimination Index, DI):	an index of how well a test item distinguishes between or among students at different levels of ability.
He/him/his:	refers to all human beings without any bias to any gender.
Skills:	refer to the abilities of the learners or examinees to do the tasks of reading. The term is interchangeably used with 'abilities'.

Testing Reading Skills: The State-of-The-Art

To appropriately gauge a test of reading skills, it is essential to discuss the nature, process, level, and types of reading. It appears equally essential to identify factors affecting the difficulty and the techniques of testing reading abilities in a foreign language.

In order to test the construct – the ability we wish to test – we need to explicitly understand what the construct is. This might include what it means to read and understand at a certain level. Yet, explicitly understanding the construct of reading skills is a taxing task.

In an academic circumstance, however, if we wait until we have a perfect understanding of our constructs before we begin to devise assessment or testing instruments, then we will never begin test construction at all. On the other hand, to refuse to get involved in designing instruments would be irresponsible and risk the danger that others, with a lesser understanding of what is involved in reading, might design the instrument to measure the reading ability. Therefore, testers have to get involved in the construction, even though they know in advance that their understanding of phenomenon – the construct – is faulty, partial and possibly never perfectible, as Alderson (2000) argues.

Theoretical framework

This research is based on these assumptions: by testing, it is possible to positively affect teaching, provided we research what we design, and we can contribute to the ever-growing body of literature of testing and teaching reading. The following model depicts the theoretical framework of the research in a very general sense.

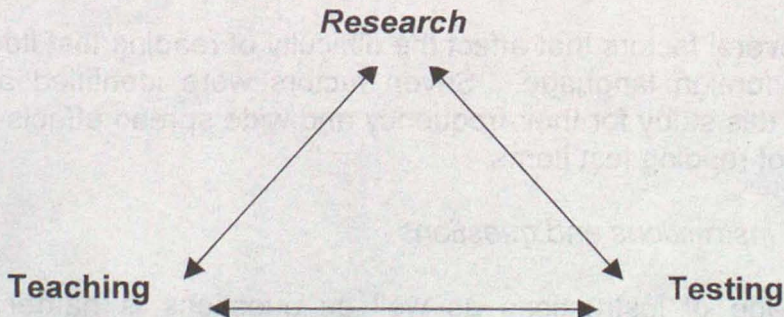


Figure 1. Teaching- research-testing model

The model reveals a triangular interrelationship between testing, research and teaching the reading abilities. It can be observed from the model that teaching methods affect testing methods and the effect can be mediated by research. The double arrows indicate the bi-directionality of the effect of each activity, i.e. teaching reading abilities is geared to testing reading abilities and research on testing informs teaching as well as testing reading.

Testing Reading in a Foreign/Second Language

For a very long time, many foreign/second language teachers have believed that poor second language reading is due to lack of good reading skills or habits in the first language. Alderson (1984) asserted that there is likely to be a language threshold beyond which second language readers have to progress before first language reading abilities can transfer to the second language situation.

In his recent work, Alderson (2000) revised his original work and acknowledged the importance of both factors – foreign/second language knowledge and first language reading skills. He commended that in second language reading, knowledge of the second/foreign language is a more important factor than first language reading skills.

Factors affecting the difficulty of reading test items

There are several factors that affect the difficulty of reading test items in a second/foreign language. Seven factors were identified and discussed in this study for their frequency and wide spread effects on the difficulty of reading test items.

Language of instructions and questions

If the language of instructions as well as questions is harder to understand than the passages themselves, the reader is presented with additional layers of difficulty. As a result, we cannot tell whether

performance is due to the difficulty of passage, language of instruction, or the type of the questions.

Types of questions

Pearson and Johnson (1978) identified three types of questions in their research on the testing of reading abilities.

- *Textually explicit questions or local comprehension questions:* This type of question requires the respondents to focus on a reading line, i.e. the information and the correct answer are found in the reading passage very easily. Some times it is termed as referential question;
- *Textually implicit or global comprehension questions:* This requires the respondents to combine information across sentences; and
- *Script-based or scriptally implicit questions:* This requires readers to integrate text information with their background knowledge since the correct answer is not explicitly found in the text itself.

Davy and Lasasso (1984) found that their textually explicit items were significantly easier than the textually implicit ones. However, Alderson et al. (1995) used this classification scheme, but did not report relationships with item difficulty or discrimination. On the whole, question types affect the reading skills test.

Text Length

Enginer (1977) argued that when texts longer than 1000 words were used, the abilities that could be measured changed. The argument is that a longer text allows testers to assess most study-related abilities that might tap syntactic and lexical knowledge. Different studies reveal that it is likely to be much easier to measure reading abilities using longer texts than by a number of short passages with associated questions.

Using longer texts to test for academic purposes is justified because it reflects the situation where students have to read and study long texts. The practice of International English Language Testing System (IELTS) in contrast with that of Test of English as a foreign Language (TOEFL), where shorter passages are being used, substantiates the claim.

Focusing on Grammar in Reading Test

Some times reading tests focus on grammar more than on reading skills. For instance, Bachman et al. (1989) found that aspects of test could account for almost 70% of variance on the item difficulty of the TOEFL reading sub-tests related to grammar more than to the academic and topical content of reading items. This is much more than would be expected of a test of academic reading ability. Alderson (1993) reported in his study conducted as part of the development of IELTS that there was high correlation between a communicative grammar test and a test of academic reading ability. It appeared that sometimes the grammar test was more closely related to a test of academic reading than the reading test was related to another parallel reading test.

Focusing on Vocabulary in Reading Test

Research consistently shows that tests of vocabulary are highly predictive of performance on tests of reading comprehension. In short, vocabulary plays a very significant role in reading test. For instance, in studies of readability, most indices of vocabulary difficulties account for about 80% of the predictive variance as Alderson (2000) reported. Clearly, vocabulary is important to text comprehension and, thus to test performance. However, it is advisable to be explicit about what it means to know and use a word in order to derive useful insights for reading test development.

Background Knowledge and Subject Knowledge

A number of studies (Carver, 1992a; Clapham, 1996; Alderson, 2000;) appear to indicate that background knowledge or subject knowledge of the candidate could affect his performance in the reading abilities test. This in turn might affect the interpretation of test results and test construction.

Lack of Awareness of Test Development in Test Construction

It is crucial that test constructors are well aware of principles and practice of test development. Alderson *et al.* (1995) presented the following top nine priorities as a model for test development.

- Identifying test purpose;
- Developing test specification;
- Guidelines for and training of item/task writers and moderation of the products;
- Pre-testing, analysis of results and revision of tests;
- Training examiners and administrators;
- Monitoring examiner reliability;
- Reporting scores and setting pass marks or cut off points;
- Test validation;
- Developing and improving tests.

Techniques For Testing Reading Abilities

It is noteworthy that there is no one best method for testing reading in a second/foreign language. There are a number of techniques or

methods for testing reading abilities. However, I will concentrate on a few ones, which are most commonly employed in different settings.

Multiple-choice techniques

Multiple-choice questions are a common device for testing the comprehension abilities of students. Some of their positive aspects are that they:

- allow testers to control the range of possible answers for comprehension;
- control, to some extent, the students' thought processes when students attempt to respond;
- create a complete marker reliability.

Despite the positive aspects, multiple-choice questions are contentious for:

- they present students with unanticipated distracters; this deliberate tricking of students might result in false measure of understanding;
- some researchers argue that the ability to answer multiple-choice questions is a separate ability, different from the reading ability;
- students can learn how to answer multiple-choice questions by eliminating unlikely distracters;
- they might be gender biased (it is speculated that male students do better in multiple choice questions than their female counter parts);
- the construction of good multiple-choice questions requires more time and expertise.

Matching techniques

In matching techniques there are two sets of stimuli: column 'A' and column 'B', which have to be matched against each answer.

Matching is subject to the same criticism of multiple-choice types in that candidates may be distracted by choices they would not otherwise have considered.

Ordering Techniques

In an ordering task, candidates are given a set of words, phrases, sentences, paragraphs, or texts and have to put them to their correct order. They seem to be superficially attractive since they offer the possibility of testing the ability to detect cohesion, overall organisation or complex grammar. However, such tasks are remarkably difficult to construct satisfactorily. The most controversial aspects of this technique are that:

- There could be more than one possible ordering sequence;
- Partially correct answers pose problems; for example, if a student gets three, or four, or five, or six elements out of ten in correct sequence, how is such a response to be weighted?

Dichotomous Items

Dichotomous items refer to test items with only two options: true/false; yes/no; agree/disagree and the like. The technique is popular apparently because of ease of construction.

The problem, of course, is that students have 50% chance of getting the answer right by guessing alone. To counteract such defects, including a third category, for instance, 'not given' or 'the text does not say' or requiring examinees to 'provide evidence' might be worth considering. However, with items intending to test the ability to infer meaning, this can cause confusion.

Editing Tests

These tests consist of passages in which errors have been introduced, which the candidate has to identify. These errors can be in multiple-choice format or in an open-ended fashion, e.g. by asking candidates to identify one error per line of a text and to write the correct answer on the opposite line. The nature of the error will determine to a large extent whether the item is testing the ability to read or a more restricted linguistic ability. Such tasks would be similar to proof-reading tasks, which could often be the 'real life' justification for editing tasks more generally.

Short Answer Tests

A semi-objective alternative to multiple choice is the short-answer question. In this method, the test-takers are simply asked questions, which require a brief response. Unlike in multiple-choice question items, it is possible to interpret the students' response to see if they have really understood the text and the question.

Short-answer questions are not easy to construct. The questions must be worded in such a way that all possible answers are foreseeable; otherwise, markers might have different possible answers.

The Free-Recall Tests

Free-recall tests require students to read a text first, then put it one side, and then write down everything they can remember from the text.

This technique is often held to provide a purer measure of comprehension, since test questions do not intervene between the reader and the text; it is believed to provide picture of the learning process of the learner. For instance, it is possible to notice from the

text how the readers have reconstructed the meaning or how they organised and restored the information.

However, some people oppose the idea of free-recall test for it tests more of memory than understanding. Unless instructions are written explicitly and clearly, and how the test is to be evaluated, marking might be problematic.

The summary test

A more familiar variant of the free recall test is the summary test. In this technique the test-takers are expected to:

- read the whole or part of the text and understand the main ideas;
- summarise these main ideas separating them from the irrelevant ones;
- organise these coherently.

Though familiar and relevant, this technique presents problems for scoring. One solution of the contamination of reading with writing is to present a number of multiple-choice summaries where the reader selects the best summary.

The gapped summary techniques

The gapped summary can be one way of overcoming the subjectivity problem experienced in marking summary test. The gapped summary technique requires students to:

- read a text, and then read a summary of the same text;
- restore the deleted words from the text; in other words, complete the gapped summary choosing possible words and phrases from a bank, e.g. box, etc., provided by the constructors.

Alderson et al (1995) conclude that such tests are difficult to write, and need much pre-testing, but can eventually work well and are easier to mark.

Information-transfer techniques

These techniques are fairly common in testing and teaching reading abilities in academic contexts in both first and second languages. The students' task is to identify in the target text the required information and then to transfer it, often on to a table, map or flow-chart. Often the answers consist of names and numbers and can be marked objectively; otherwise, they require phrases or short sentences and need to be marked subjectively.

Two problems related to this technique are in order:

- They may be cognitively or culturally biased. In some cultures, use of tables, charts or tables may be uncommon. This might threaten the test-taker;
- The tasks can be very complicated. It may take more time for the testee to decide what should go where in the table or map. In other words, the information transfer technique adds an element of difficulty that is not in the text of the test.

What is common to all the techniques discussed here is that they fail to measure the construct they are supposed to measure. Therefore, the best method, if at all there is one, is to use different techniques to test the academic reading abilities.

Methodology

This section explains the methods used in carrying out the study by giving special emphasis to the analysis of data. It should be noted at the outset that the methodology involves both qualitative and quantitative data. As this is a preliminary study, it might be

considered as an evolving research that would take definite shape as it progresses.

The Research Context

The study took place in the Department of Foreign Languages and Literature, Addis Ababa University. It focused on the reading ability test of College English, designed by the testing committee of the Department.

The Research Participants

As was pointed out in the introduction part, the study involved instructors of the Department directly, and students that took the course between 1997 and 2001 indirectly.

Instruments and Procedures Used in Data Collection

Questionnaires were used to investigate the instructors' response concerning the reading ability test. In order to assess the instructors' views, questionnaires containing four themes were designed; namely, the clarity of instruction, communicative content, format, and the overall assessment of the test. The questionnaires were distributed to 15 instructors out of 37, who taught the course, College English. All of them responded to the questionnaire.

The rationale for using the questionnaire in this study is that it is handy and more familiar means of gathering data from instructors, who have little time to spare.

The responses were computed using Statistical Package for Social Scientists (SPSS) and the analysis was divided into four categories. These are instructors' response to:

- the clarity of instructions and questions in the test
- contents of passages in the test

- format of the test
- overall assessment of the test.

50 examination scripts, marked by the instructors, were used to determine the level of difficulty and index of discrimination of the reading test items. To this end, Item Analysis (ITEMAN), version 3.5 was employed.

Reading skills testing techniques of College English of 1997 – 2001 were re-examined in the light of existing literature.

Data analysis

The data were analysed using several strategies. First, the data captured through questionnaire were categorised. Then, the raw data were reduced and SPSS was run. Only selected results were displayed in tables in the result part of the study; the remaining tabulated results were annexed in Appendix B. Item analysis raw data were reduced and the ITEMAN was run. Very limited and specific question types were analysed, and the tabulated items were in Appendix A.

Limitation of the method

- As this is a preliminary study, subjects of the study were limited.
- It employed predominantly the quantitative methods; the conclusion drawn may also be limited consequently.

Result

Focusing on item analysis, the study reported here examined whether the reading test items measure the academic reading abilities of the college students. It also investigated instructors' responses and the techniques used in the test of reading abilities.

Item analysis

Among four types of College English test items, given as one integrated test in the final exam of a semester, this research focused on the reading skills/abilities test. The summary of each item is given below:

Table 1: Summary of Types of the College English Test

Type of test	Writing	Grammar	Reading Skills	Vocabulary
Type of scale	Subjective	Dichotomous*	Dichotomous*	Dichotomous*
Number of items	5	20	35	15
No. of examinees	50	50	50	50

*For ITEMAN purpose, multiple-choice, true/false, short answer, matching, gapped summary and information transfer questions of 1998/1999 College English reading skills test were given numbers and treated as dichotomous.

To avoid complications, I took the first 5 out of 35 items of the reading skills test to examine (a) discrimination index and (b) level of difficulty or facility value (see Appendix A for the detail).

Table 2: Item Analysis of the First 5 Sample Questions of the Reading Skills Test

Item Statistics				Alternative Statistics			
% correct	Disc. Index	Facility value	Alternative	% correct	Endorsing		Facility Value
					Low	High	
68	.43	.41	1	8	14	0	-.18
			2	18	29	7	-.27
			3	6	7	0	-.16
			4	68	50	93	.41*
60	.43	.31	1	40	64	29	-.25
			2	60	36	71	.31*
42	-.07	-.06	1	58	50	57	.11*
			2	42	50	43	-.06
54	.07	.10	1	54	57	64	.10*
			2	46	43	36	-.10
70	.36	.32	1	70	50	86	.32*
			2	30	50	14	-.32

* signifies the attraction power of alternatives or options in the reading test items.

• indicates point biserial or facility value.

The two measurements used to gauge the reading abilities test are the facility value (FV) and the discrimination index (DI). The facility value measures the level of difficulty of an item, and the discrimination index measures how well a test item distinguishes between or among students at different levels of ability. In some books, discrimination index is defined as the extent to which the results of an individual item correlate with the whole test.

If the item is working well, we should expect more of the top scoring students to know the answer than the lower ones. If the strongest students get an item wrong while the weakest students get an item right, then there is clearly a problem with the item, and this needs investigating.

In an ideal situation, the highest DI possible is 1:00, which is achieved if all the students in the top group get an answer right, and none of the students in the bottom group does. Such items, however, are very unusual. An item with a DI of 0.50 is usually considered to be discriminating well. Often item writers are content with DI of 0.40 or above, but there are no rules as to what DI are acceptable, since the possibilities of getting high DI varies according to the test type and range of ability of the examinees.

An item with a lower value of FV and fairly high value of DI is still considered as discriminating well (Jackson and Messick, 1967; Alderson et al, 1995; Field, 2000). Sometimes, however, an item has a negative DI, which means that more students in the bottom group were correct than in the top group. There is obviously something wrong with such an item and it should be revised or discarded.

This study intended to measure the FV and the DI of the reading abilities test items. As noted above, 35 items of reading test of 50 examinees were taken as a sample and the first 5 questions were focused on.

As can be observed in Table 2, item 1 of the test is multiple-choice comprising 4 alternatives. The DI is 0.43, which is below 0.50 and FV is 0.41. It should be noted that in the above table there are two sets of facility values. The first set in the item statistics category describes the FV of the item as a whole, whereas the second FV indicates the difficulty level of each alternative. In the latter case, we can see how many percent of the lower and higher group are attracted to each alternative answer.

Alternative 1 of item 1 of the reading test attracted 14% of lower group students and none from high scoring group. In other words this option was so weak that it failed to attract top group students; hence that option was a loss. And its FV is negative i.e. -.18 indicating that something was wrong with the distracter. Distracter 2 could attract 29 % of lower group and only 7 % of higher group and still with negative FV implying the same shortcoming in the item. The same is true with distracter 3. However, the right answer, i.e. option 4 with FV of 0.41, was able to attract 93 % of the high group and 50 % of the low group.

Item 2 in the same table has the discrimination index of .43 with difficulty level of .31. It could be seen from the table that higher number of the high achieving students than low achieving ones were attracted to the right answer suggesting that the item, with some modification and piloting, would be a useful item to measure the academic reading skills of the college students.

The third item with overall negative discrimination index and level of difficulty could have been discarded from the out set if there were pre-testing sessions; it failed to assess the students capacity to read. As can be observed in the table, a good number, 57 % of high group students, went for the wrong answer.

In item 4, in spite of very low discrimination index and difficulty level, it is apparent that the majority of high group students were attracted to the right answer, i.e. 64 %. The item might have been of high DI and FV if it were pre-tested.

Likewise, item 5 appears to discriminate to some extent the high and low achieving students; consequently, 84 % of high group went for the right answer. The item still needs piloting because its discrimination power and level of difficulty are below the standard. (For details see Appendix A).

Questionnaire analysis

Instructors' responses to items of the reading test were portrayed in the following tables. Clarity of instruction and questions of the test; content of the test; format of the test; and the overall assessment of the test were the items analysed. For the sake of clarity, I have taken only one question from each theme, and annexed the rest in Appendix B.

Table 3: Instructors' response to clarity of questions and instructions in the Test

Students are informed in the instruction as to how to read for specific reading skills such as skimming or scanning.	No.	%
Strongly agree	4	26.7
Agree	6	40.0
Disagree	3	30.0
Strongly disagree	2	13.3
Total	15	100

Table 3 provides instructors' responses regarding the specificity of instructions in the reading part of the College English. As can be seen in the table, about 67 % or 10 out of 15 of them positively responded. In their view students were provided with clear instructions as to how to read for different skills of reading. However, 5 out of 15 or nearly 43 % responded negatively to the question. They did not give any explanation why they negatively reacted to the question. In the literature it was noted that the language of instructions as well as questions in the test should be easier than the actual test or the passage.

Table 4: Instructors' Response to the Specific Skills of Reading Test

The test measures the skills specified in the instructional objectives of the course of the course.	No.	%
Strongly agree	1	6.70
Agree	9	60.0
Disagree	4	26.7
Strongly disagree	1	6.70
Total	15	100

As can be observed in Table 4, almost 67 % of instructors or 10 out of 15 indicated that the reading skills test measured the skills specified in the instructional objectives of the course. Nonetheless, the data did not indicate which types of skills were specified in the test. 5 instructors i.e. over 30 % were doubtful about the reading skills test measuring the skills specified in both instructor's manual and students' text book. Though the instructors were not given the opportunity to state justifications for their doubt, one can speculate that the lack of piloting of the tests, which might rectify any defective items in the test, could be one possible explanation.

Table 5: Instructors' Responses to the Format Of The Reading Test

Could you please suggest alternative format for testing vocabulary in the reading skills/abilities test	No.	%
Gap filling	1	6.7
Sentence construction	9	60.0
Asking students to find opposite, similar and collocating words	3	20.0
Gapped summary	1	6.7

Table 5 summarises the instructors' responses concerning alternative formats to the matching layout in testing vocabulary in the reading skills test. As can be seen in the table, more than half, i.e. 60 % of the instructors recommended that requiring the examinees to construct sentences with the highlighted words in the passages might reveal their level of understanding. About 40 % of the instructors suggested

various ways of testing vocabulary in reading skills test. Among them are making students use different vocabulary learning strategies; letting them fill gapped summary; requiring them to write definitions for the words selected from the passage; and asking them fill gaps.

Table 6: Instructors' Responses To The Overall Assessment of The College English Reading Skills Test

What is your overall assessment of the College English Reading Skills Test?	No.	%
The reading test items are up to the standard.	9	60.0
Passages are long and boring.	2	13.3
The reading test needs overall improvement.	1	6.7
Texts are difficult.	1	6.7
Texts, i.e. passages are long for the freshman level.	1	6.7
The passages are biased towards science	1	6.7
Total	15	100

Table 6 provides the overall assessment of the instructors of the reading test items of College English. The table reveals that instructors agree with the overall format of the present testing practice; 60 % of them favoured the test as it stands now.

About 13 % articulated that the passages are long and boring. They expressed their dissatisfaction with the content as well as the length of the passage. However, research findings elsewhere in academic contexts recommend and prefer long and authentic passages to short and unrealistic ones. The justification is that long passages, as long as 1000 words in an academic reading test, reflect the reality of academic settings, where students are required to read even longer texts than that.

Nearly 28 % of the respondents gave different views regarding the overall assessment of the test: the reading test needs overall improvement; texts are difficult, long and biased towards science. They appear to imply that test constructors select passages, which might unfortunately favour some of the students. A clear signal is that

there should be a mechanism to balance passages so that there may be less bias towards one discipline.

Techniques of testing the reading skills of College English

To assess consistency of the Department's current practice of testing reading with the state-of-the-art of testing reading, I have selected College English exam scripts of 1997/8 – 2000/01. The following techniques were found to be constantly employed in the test over the academic years mentioned:

- Short-notes (short answers);
- Gapped summary;
- Identifying paragraph functions (writing paragraph summaries);
- Multiple-choice questions;
- True/false with the third category – providing evidence for the answers;
- Matching of words based on the context (guessing in a context)
- Matching of summaries of paragraphs with their numbers in the passage;
- Completion of statements;
- Reference;
- Inference
- Information transfer.

Seen in the light of the discussion in the literature part, it is apparent that the reading test of College English, to a degree, is consistent with the recent developments in testing reading abilities. The point of

concern, however, is that these same techniques have remained the same for a long time; they have not been assessed for their effects on test interpretations and test construction.

Discussion

As stated in the first part, this study raised two questions:

- Do the items of reading skills test of College English measure the academic reading abilities of the students?
- How do instructors perceive the content of passages, format, and specific items of the reading abilities test of College English?

Item analysis was carried out to determine the index of discrimination and the level of difficulty of the test items. Questionnaires were employed to find out the reactions of the instructors. And testing techniques of academic reading abilities of 1997/8 – 2000/01 were revisited to assess the types of techniques in the light of the current research on testing academic reading abilities.

The overall result of item analysis (see appendix A) indicates that the test items, to a degree, measure the academic reading abilities of college students. However, it should be noted that this preliminary study has examined only 35 items of 50 examinees. Therefore, the results of this study cannot be taken as conclusive evidence for all the reading test items of College English to be testing the construct intended to be tested. It only offers some insight into the reading skills test construction, research on testing, and teaching academic reading skills at first year level at AAU.

As can be observed in Appendix A, about 20 items out of 35 in the reading test fall short of coming up to the facility value of 50% and discrimination index of 0.50. The main reasons for most items to be very weak at discriminating index and difficulty level appear to be the

lack of test-specifications to guide the test constructors and piloting of the items to check:

- clarity of instructions and test questions;
- types of techniques, i.e. the format being used for test construction;
- the kind of language being elicited in open-ended questions such as summarising;
- reading constructs that should be tested;
- accuracy and comprehensiveness of answer keys;
- the content of reading passages;;
- time students need for completing the test.

Long passages in the reading abilities test of College English have been the points of criticism. However, recent studies on academic reading abilities tests seem to support the principles being applied by the testing committee of the Department. The justifications are that longer texts reflect real life reading practice the students do on their academic reading in the main streams. This seems to be another fairly strong point of the College English reading skills test. Moreover, existing research tends to suggest that test constructors should worry more about the communicative content of a passage than its length.

By and large, the reading skills test of College English is consonant with the overall objectives of the course; however, questionnaire responses, item analysis of the test, and the review of current testing practice of College English, imply that significant improvements and further research on the reading skills test are imperative.

Implications for Testing And Teaching Reading In A Foreign Language Context

From the above discussion, it is possible to draw the following implications for the testing and teaching of reading skills in a foreign language context.

- Test designers need to consider to what extent their tests reflect and build upon what recent research into reading suggests about the process, not just the product;
- Students should be encouraged to read longer texts; it should be possible for them to do the tasks, either in the test or in the classroom reading, in the time available without discouraging students because of their difficulty level;
- Test designers should be provided with test specifications and test development processes with up-to-date test construction techniques;
 - Test items should be pre-tested to make sure that the test is fool proof. No matter how well designed an examination may be, and however carefully it has been edited, it is not possible to know how it will work until it has been tried out by students, as a research report by Alderson, Clapham and Wall (1995) confirms;
 - Apart from calculating the facility value and discrimination index, the future item analysis, in large-scale research, should focus on the students' wrong answers. These could throw light on how the students understood the task and whether the item is testing what is intended. They might also reveal inaccuracies and omissions in the key, and will highlight ambiguities in the marking scheme so that they could help to measure the measurement more accurately.

References

- Alderson, J. C. (1984). *Reading in a Foreign Language: A Reading Problem or a Language Problem?* In J.C. Alderson and A. H. Urquhart(eds.), **Reading in a Foreign Language**. London: Longman.
- _____ (1993). *The relationship between grammar and reading in an English for academic purposes test battery*. In D. Douglas and C. Chapell (eds.), *A new decade of language testing research: Selected papers from the 1990 :Language Testing Research Colloquium*. Alexandria, VA: TESOL.
- _____ (2000). **Assessing reading**. Cambridge: Cambridge University Press.
- _____, Clapham, C., and Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- _____, and Lukmani, Y. (1989). *Cognition and Reading: Cognitive Levels as Embodied in Test Questions*. **Reading in a Foreign Language** 5 (2), 253 – 270.
- Bachman, L.F., et.al. (1989). *Content Analysis and Statistical Modelling of EFL Proficiency Tests*. Paper presented at The 11th Annual Language Testing Research Colloquium, San Antonio, Texas.
- Carroll, P. L. Devine, J. and Eskey, D. (eds.) (1988). **Interactive Approaches to Second Language Reading**. Cambridge: Cambridge University Press.
- Carver, (1992a). *Effects of prediction activities, prior knowledge, and text type upon amount comprehended: using reading theory to critique schema theory research*. **Reading Research Quarterly**.
- Clapham, C. M. (1996). *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension*. Cambridge: Cambridge University Press.
- Davey, B.and Lasasso, C. (1984). *The interaction of reader and task factors in the assessment of reading comprehension*. **Experimental Education** 52, 199 – 206.
- Engineer, W. (1977). *Proficiency in reading English as a second language*. Unpublished PhD thesis, University of Edinburgh, in Alderson, J. C. (2000). **Assessing Reading**. Cambridge: Cambridge University Press.

- Field, A. (2000). **Discovering Statistics Using SPSS for Windows**. London: SAGE publications.
- Goodman, K. S. (1969). *Analysis of oral reading miscues: Applied psycholinguistics*. **Reading Research Quarterly** 5, 9 – 30.
- Goodman, K. S. (1982). **Process, theory, research**. Vol.1. London: Routledge and Kegan Paul.
- Halliday, M. A. K. (1979). **Language as social semiotic**. London: Edward Arnold.
- Jackson, D.N. and Messick, S. (1967). **Problems in Human Assessment**. New York: McGraw-Hill Book Company.
- Macnamara, T. (2000). **Language Testing**. Oxford: Oxford University Press.
- Munby (1978). **Communicative Syllabus Design**. Cambridge: Cambridge University Press.
- Pearson, P. D. and Johnson, D. D. (1978). **Teaching Reading Comprehension**. New York, J: Holt, Rinehart and Winston.
- Smith, F. (1971). **Understanding Reading**. New York, NY: Holt, Rinehart and Winston.
- Smith, F. (1980). **Reading**. New York, NY: Holt, Rinehart and Winston.
- Widdowson, H.G. (1979). **Explorations in Applied Linguistics**. Oxford: Oxford University Press.

Appendix A

Item analysis of the questions of the reading skills test of College English

Item statistics				Alternative statistics			
% correct	Disc.	Point.	alter	%correct	Endorsing		Point
					Low	High	
68	.43	.41	1	8	14	0	-.18
			2	18	29	7	-.27
			3	6	7	0	-.16
			4	68	50	93	.41*
60	.43	.31	1	40	64	29	-.25
			2	60	36	71	.31*
42	-.07	-.06	1	58	50	57	.11*
			2	42	50	43	-.06
54	.07	.10	1	54	57	64	.10*
			2	46	43	36	-.10
70	.36	.32	1	70	50	86	.32*
			2	30	50	14	-.32
46	.50	.31	1	46	29	79	.31*
			2	54	71	21	-.31
36	.71	.62	1	36	14	86	.62*
			2	64	86	14	-.62
20	.50	.50	1	20	0	50	.50*
			2	80	100	50	-.50
52	-.07	.02	1	52	57	50	.02*
			2	48	43	50	-.02
6	.07	.12	1	6	0	7	.12*
			2	94	100	93	-.12
30	.29	.34	1	30	29	57	.34*
			2	70	71	43	-.34
60	.57	.53	1	60	21	79	.53*
			2	40	79	21	-.53
54	.43	.40	1	54	36	79	.40*
			2	45	64	21	.40
36	.50	.39	1	36	21	71	.39*
			2	64	79	29	-.39
66	.36	.34	1	66	50	86	.34*
			2	34	50	14	-.34

Item statistics				Alternative statistics			
% correct	Disc.	Point.	alter	%correct	Endorsing		Point
					Low	High	
50	.71	.49	1	50	14	86	.49*
			2	50	86	14	-.49
20	.07	.12	1	20	14	21	.12*
			2	80	86	79	-.12
4	.14	.33	1	4	0	14	.33*
			2	96	100	86	-.33
22	.21	.21	1	22	14	36	.21*
			2	78	86	64	-.21
40	.64	.41	1	40	14	79	.41*
			2	60	86	21	-.41
26	.43	.37	1	24	7	50	.37*
			2	74	93	50	-.37
34	.64	.55	1	34	7	71	.55*
			2	66	93	29	-.55
34	.64	.59	1	34	7	71	.59*
			2	66	93	29	-.59
46	.71	.59	1	46	14	86	.59*
			2	54	86	14	-.59
36	.71	.67	1	36	7	79	.67*
			2	64	93	21	-.67
72	.50	.41	1	72	43	93	.41*
			2	28	57	7	-.41
64	.29	.35	1	64	50	79	.35*
			2	36	50	21	-.35
68	.71	.61	1	68	21	93	.61*
			2	32	70	7	-.61
36	.57	.45	1	36	7	64	.45*
			2	64	93	36	-.45
78	.07	.06	1	78	64	71	.06*
			2	22	36	29	-.06
54	.71	.53	1	54	14	86	.53*
			2	46	86	14	-.53
36	.57	.53	1	36	21	79	.53*
			2	64	79	21	-.53
42	.43	.29	1	42	14	57	.29*
			2	58	86	43	-.29
54	.07	.04	1	54	43	50	.04*
			2	46	57	50	-.04
54	.14	.11	1	54	29	43	.11*
			2	46	71	57	-.11

Appendix B

Table 2: Instructors' Responses to Questions of Clarity of Instructions in the Test

	Response	
	No.	%
Instructions For Each Item Are Clear		
Strongly agree	8	53.3
Agree	5	33.3
Disagree	1	6.7
Strongly disagree	1	6.7
Total	15	100
Students Are Informed In The Instruction As To How To Read For Specific Reading Skills Such As Skimming Or Scanning	No.	%
Strongly agree	4	26.7
Agree	6	40.0
Disagree	3	30.0
Strongly disagree	2	13.3
Total	15	100
Words Used In The Instructions Are Easier Than The Words Used In The Test	No.	%
Strongly agree	3	20.0
Agree	9	60.0
Disagree	1	6.7
Strongly disagree	2	13.3
Total	15	100
There are clear examples showing how students should answer the items/tasks in the test	NO.	%
Strongly agree	1	6.7
Agree	6	40.0
Disagree	6	40.0
Strongly disagree	2	13.3
Total	15	100

Table 3: Instructors' Response To The Content Of Reading The Test

The test measures specified skills	Response	
	No.	%
Strongly agree	1	6.7
Agree	9	60.0
Disagree	4	26.7
Strongly disagree	1	6.7
Total	15	100
The test assesses skills, which are not taught/learned in the college English course	No.	%
Strongly agree	2	13.3
Agree	7	46.7
Disagree	3	20.0
Strongly disagree	3	20.0
Total	15	100
The tasks/items in the test make students use the reading sub-skills, i.e. skimming, Scanning, Referencing, And So on	No.	%
Strongly agree	2	13.3
Agree	9	60.0
Disagree	2	13.3
Strongly disagree	2	13.3
Total	15	100
The content of the test covers the communicative response ability of reading	No.	%
Agree	8	53.0
Disagree	5	33.3
Strongly disagree	2	13.3
Total	15	100
The content of the passages in the test are fair for test-takers with different backgrounds, e.g. gender, ethnicity, religion, physical impairment, field of study, and the like.	No.	%
Agree	8	53.3
Disagree	2	13.3
Strongly disagree	5	33.3
Total	15	100

The Test Measures Specified Skills	Response	
There are insensitive sections or language items in the test, e.g. offensive, obscene and taboo language items.	N0.	%
Agree	4	26.7
Disagree	11	73.3
Total	15	100
The length of the reading passage in the test is appropriate to the level of the test-takers	N0.	%
Strongly agree	1	6.7
Agree	7	46.7
Disagree	6	40.0
Strongly disagree	1	6.7
Total	15	100

Table 4: Instructors' Responses To The Format of The Reading Test

There a suitable number of questions in the test	Response	
	No.	%
Strongly agree	2	13.3
Agree	9	60.0
Disagree	4	26.7
Total	15	100
There are some possible answers or questions, which unfortunately help the candidate to answer other questions in the test	No.	%
Strongly agree	1	6.7
Agree	5	33.3
Disagree	8	53.3
Strongly disagree	1	6.7
Total	15	100
There are items/tasks that are too difficult for the candidates to answer	No.	%
Strongly agree	1	6.7
Agree	2	13.3
Disagree	8	53.3
Strongly disagree	1	6.7
Total	12	80
<i>Missing -not answered</i>	3	20.0
Total	15	100
There are items that are too easy for the students to answer	No.	%
Agree	2	13.3
Disagree	7	46.7
Strongly disagree	2	13.3
Total	11	73.3
Missing (unanswered)	4	26.7
Total	15	100

Questions In The Test Are Sequentially Ordered	Response	
	No.	%
Strongly agree	2	13.3
Agree	9	60.0
Disagree	4	26.7
Total	15	100
The Only Appropriate Item Response Formats For Testing Vocabulary In The Reading Test Are Matching And Multiple- Choice Questions	No.	%
Strongly agree	1	6.7
Agree	5	33.3
Disagree	8	53.3
Strongly disagree	1	6.7
Total	15	100
Could You Please Suggest Alternative Means For Testing Vocabulary In The Reading Skills/Abilities Test	No.	%
Gap filling	1	6.7
Sentence construction	9	60.0
Learning strategies	3	20.0
Banked cloze	1	6.7
Asking the students to write the definition words	1	6.7
Total	15	100
The timing of the reading skills test is realistic.	No.	%
Strongly agree	1	6.7
Agree	9	60.0
Disagree	3	20.0
Strongly disagree	1	6.7
Missing	1	6.7
Total	15	100