Appraisal of the Extent Tanzanian EFL Testing Complies to Ideals of Test Construction: A Case of 2012 Certificate of Secondary Education English Language Examination

Zelda Elisifa

Department of Linguistics and Literary Studies

The Open University of Tanzania

Kawawa Road, Kinondoni Municipality

P.O Box 23409, Dar es Salaam

Tanzania

Tel: 255-754-258700   E-mail: elisifazem@yahoo.co.uk

Abstract

This study seeks to show the extent to which EFL testing at Certificate of Secondary Education Examination level is at par with ideals of examination construction, notably in the area of content areas. To achieve this content analysis was employed to 2012 paper as a single case study. The major aspects in are those given by Harris (1969;) namely determining course objectives, dividing the general course objectives into their components, the time factor for the examination and the speediness, and writing examination directions.

The findings revealed that the testing content is more biased towards literate skills of writing and reading at the expense of oral skills, even though the objectives include speaking skills. The Skill that is totally omitted, both in the instructional objectives and examination content areas, is listening. It is recommended that more language experts be involved in all procedures of examination setting.

Keywords: Testing, Foreign Language, Test Fidelity, Table of Specification

1. Introduction

A test is essentially a collection of items the design and construction of basically follow the principles of discrimination, score variance, reliability, and validity. A good item construction process is one that enhances the discrimination power, score variance, reliability and evidence of validity for the intended interpretation and use of scores from the overall test (Suen & McClellan, 2003). Any construction of a language test should thus involve planning the preparing test items and directions, submitting the test to review and revising on the basis of review, pre-testing the material and analyzing the results, assembling the form of the test and finally reproducing the test (Harris, 1969). This is because test might be used to obtain information to guide the placement of students into appropriate classes or programs.

According to Salisbury (2005), there are three phases in the work of the test writing: First is exploratory Phase in which one searches through possible texts, or, possibly, contexts; second, called concerted Phase, is concerned with working in an intensive and concentrated way to prepare text and items for first submission, and, third, refining phase, is about polishing/improving the test paper in an effort to make it conform more closely to domain requirements. This is after either self-, peer- or editor-review. At the same time, Khalifa and Weir (2009) suggest that question paper production perhaps depends as much on the shared expertise and values of the item production team as on the procedures set out in item writer guidelines.

British Council (n.d.) lists stages of test construction, the first three stages of which are *commissioning, pre-editing* and *editing* all of which involve gathering and choosing appropriate test content that reflects the aims of the Academic and General Training modules. Once the best material has been selected, it is then given to representative groups of language learners to check that each question – or *item* – is at an appropriate difficulty level for the learners, a stage known as *pretesting*. Approved material is stored in an *item bank* and can then be introduced to *live* tests – tests that are used as the basis for awarding official IELTS certificates – through a process known as *standards fixing*.

Among the factors for language test construction include a) the difficulty of complete test versions and the range of difficulty of individual items, b) the balance of topic and genre, c) the balance of item format (i.e. the relative number of multiple choice and other item types across versions), and d) the range of listening/reading skills tested (British Council, n.d.).

However, there are practices by some language examiners to narrow the scope of testing areas. However, Ross (2005) cautions that narrow testing practices have undesirable influences on teaching, citing an example of tests which encourage memorizing facts rather than understanding relationships lead to teaching which ignores understanding regardless of national needs for people with such skills. One possible consequence, he argues, is for education authorities (and the general public) to lose confidence in the examination system because individuals who depend on more relevant items may find it incomprehensive and inadequate tool of language assessment.

The present study is delimited to the notion of test content analysis, which provides a summary of the intentions of the curriculum expressed in content terms. It seeks to answer these questions: Which content is supposed to be covered in the curriculum? Are there significant sections of this content? Are there significant sub-divisions within any of the sections? Which of these content areas should a representative test include? (Ross, 2005). Rosss (Ibid) refers to this as a test blueprint, which he defines as a specification of what the test should cover rather than a description of what the curriculum covers. This test blueprint, he posits, should include the test title, the fundamental purpose of the test, the aspects of the curriculum covered by the test, an indication of the students for whom the test will be used, the types of task that will be used in the test (and how these tasks will fit in with other relevant evidence to be collected), the uses to be made of the evidence provided by the test, the conditions under which the test will be given (time, place, who will administer the test, who will score the responses, how accuracy of scoring will be checked, whether students will be able to consult books (or use calculators) while attempting the test, and any precautions to ensure that the responses are only the work of the student attempting the test), and the balance of the questions. Additionally, adds Ross, the time factor is of crucial importance: How much time will students have to do the actual test? What time will be set aside to give instructions to those students attempting the test? Will the final number of items be too large for the test to be given in a single session? When this happens (as has been the case in NCSEE in Tanzania in the name of 'standardizing') the authorities often increase the score for a pass hoping that this will help regain public confidence. However, requiring a higher score on an inadequate test cannot solve a problem.

At another level, argues Carroll (1973), the persistent problems in foreign language testing are considered, inter alia, to fall under four headings:

The problem of validity; that is, making sure that the measurements and assessments we obtain reflect what we want them to reflect;

The problem of scope; that is, making sure that we measure or assess all the varied components of foreign language competence and skill;

The problem of efficiency; that is, obtaining the best assessments we can obtain within the limits of time and resources available for the construction and administration of the assessments, and

The problem of how tests relate to the wider context of instruction- for example, the degree to which testing either enhances instruction or, contrariwise, distorts it through undesirable feedback effects from the tests.

Our concern here is on the aspects of scope and efficiency. Elaborating more on the scope, Caroll gives it a contextual definition to refer to the question of whether foreign language tests can provide a properly comprehensive measurement of the various skills and competencies that are embodied in the concept of 'full foreign language competence,' and whether it is possible to differentiate those skills from one another in order to provide piffles of relative attainments in different skills.

Constructing the test therefore needs enough preparation bearing in mind the objectives of the course and the purpose of the test, and therefore the items should focus on these two. This paper is an analysis of English Language CSEE paper of 2012. It will provide the evidence of the examiner's fidelity or contrary thereof in the ideals of stages test construction.

2. The Methods

Content analysis was made of a single English Language examination for Certificate of Secondary Education Examination for 2012. The aim was to delve deeper into the understanding of the extent to which a single examination is at par with the ideal steps of Harris (1969) in test construction; namely, i) test construction steps of determining course objectives, ii) dividing the general course objectives into their components, iii) the time factor for the examination and the speediness, iv) writing examination directions and, iv) the ideas of producing and reproducing the final examination as given.

3. The Findings

The findings are presented in thematic chunks as phrased out by Harris' stapes of test construction as presented and analyzed below.

3.1 Determining General Course Objectives

The general course objectives, as per 2010 English Language Syllabus for Secondary Schools Form I-IV (MOEVT, 2010), for English Language teaching for Secondary Schools are as follows:

- ➤ Express oneself in both spoken and written English;
- ➤ Read widely for pleasure and for information;
- ➤ Perform a variety of writing tasks according to their required conventions;
- ➤ Use one's knowledge of English to demonstrate awareness and consciousness of basics of society and the part one can play in its development;
- ➤ Use one's knowledge of English in furthering one's education, and
- ➤ Communicate effectively with other speakers of English both inside and outside the country.

As can be seen the objectives are inclined towards mastery of productive language skills (speaking and writing) and completely omit the listening skills in spite of its being regarded as 'integrated' language skill (with reading), since, as Carroll (1968) posits, it depends (or should depend) on a wide variety of detailed competences in particular aspects of the language--its phonology, spelling, grammar, and lexicon. What is worth noting here is that the objectives are about the learner's proficiency, which, according to Stern (1983:341), refers to the actual performance of given individual learners or groups of learners, rather on display of mastery of grammatical rules in the sense that O'Grady, Dobrovolsky and Katamba (1996) gives as the learner's mental system that allows him/her to form and interpret that words and sentences of the language.

3.2 Dividing the General Course Objectives into their Components

At this second step, the general objectives are broken into their specific components according to the constructions made by the examiner(s). As Harris suggests, learners are to be

measured in listening, speaking, reading and writing. The examination paper in the current study had four (4) sections which included comprehension and summary, patterns and vocabulary, language use and response to reading. This are elaborated as follows;

Comprehension and Summary: In this part the student was given a passage to read and there after answering the questions according to the passage read. Questions from this section were two. Question 1 had two parts, part (a) with 4 multiple choice items and part (b) with three items on which the student was supposed to complete the table according to the information read from the passage. In question two the candidate was required to provide a summary of five (5) sentences from a given passage.

Patterns and Vocabulary: In this section, several activities were required to be performed by the student. And these included; four (4) completion items, eight (8) filling the gaps, four (4) short answers from family tree, eight (8) items for writing correct sentences and four (4) rewriting sentences to the instructions given. From this section a total of 24 items were provided.

Language Use: From this section, the candidate was required to answer a total of four (4) questions. The first sub-section required the student to rearrange the sentences in a sequence so as to make a meaningful paragraph; the second sub-section wanted the student to match the items so as to make meaningful sentences. From question 8 and 9, a total of 10 items were provided. Question number 10 asked the student to write essays, invitations and letters according to the instructions given. A total of two (2) questions were to be answered from this question.

Response to Reading: Here the candidate was required to answer two questions by considering to the readings done in the course of study. These are poetry, novels, plays or short stories. A total number of two questions were to be answered from this section.

In general, the whole examination consisted of a total of 55 questions 4 of which were multiple choice items from a given paragraph, 3 completing the table, summary of 5 sentences, 4 completion questions, 8 filling the gaps, 4 short answers, 8 writing correct sentences, 4 rewriting sentences accordingly, 5 re-arranging sentences, 5 matching items, and 4 essay writing questions, which makes a total of 55 questions. In general, the components and language skills measured from this paper are summarized in the Table 1 below.

Table 1. Skills Representation in the Examination

Components	Language Skills			
	Listening	Speaking	Reading	Writing
Phonology/Orthography			√	√
Structure			√	√
Vocabulary			√	√
Rate and general fluency			√	√

The data in table 1 above are an illustration of what Powers (2013) calls comprehensive assessment of English-language skills — reading, listening, writing, and speaking, related to linguistic levels of vocabulary, syntax, and orthography. However, from Table 1 above, we also note that listening and speaking skills were not measured in this examination. This is due to tendency of languages testers focusing on non-interactive tasks (Kluitmann, 2008).

3.3 Establishing the General Design of the Test

Among the issues to be borne in mind in establishing the test of the test, observes Milanovic (2002) is how much time is available for the test. The given examination in our current study provided a maximum of three (3) hours to finish all the given questions. Here we want to show the extent to which the time provided was sufficient to do the given questions, considering the degree of speediness of students.

As it has been said, the maximum of three hours were provided for the candidates to finish the required tasks. Assuming that at least 20 minutes were for administrative procedures like seating the students and handing out materials, giving general directions, collecting the materials at the end of the testing period, handling unanticipated problems etc, 160 minutes out of 180 remain for the actual test.

When considering the specify the individual variability in writing speed of individual examinees, tone notes that some are slow in reading, understanding the question and starting answering them, so they take much time in finishing the examination as compared to others. With this case, allowance was made for some minutes for slow-exam writers, and hence a fair time allotment for the achievement of the test is discussed below.

We now have to divide those skills into specific item types and see whether the time provided was sufficient or not. The item types provided included multiple choice, matching items, completion items, filling the gaps, short answers, and essay questions from various sections.

Assuming that multiple choice questions takes 20 minutes in reading the passage and answering them, out of 150 minutes 130 remains. Then, from the same passage, the student was required to complete the table; this will be treated in short-answer item under which three questions were given. 5 minutes are provided for answering these questions. Question number two required the student to read and summarize the passage given, 15 minutes are assumed to be used to answer this question.

Another part into which the candidate was tested in this paper was vocabulary, under which 4 completion items were given, and a maximum of 5 minutes is assumed to finish them, with an assumption of 1.25 minutes per question, fill in the gaps had 8 items to be answered, an estimation of 45 seconds per each, which makes a total of 6 minutes to complete this part.

Short-answer questions were provided after reading a small passage about family tree. Here, 4 questions were set all of which were obligatory; so an estimation of 6 minutes for reading and answering the questions is provided. Writing correct sentences with 8 items was another question, where provision of 8 minutes is assumed for reading, recognizing the problem and re-writing into correct sentences.

The last part in vocabulary section required candidates to re-write the sentences according to the instructions given. 4 questions were provided, and an estimation of 5 minutes is provided with a average of 1.25 minutes per question.

Language use was another section to be answered by the student, where 5 sentences were to be re-arranged in a logical sequence, and we assume that 5 minutes were provided to answer these questions with an estimation of 1 minute per each question. Another type of item from this section was matching items under which the student was required to match the given sentences so as to make meaningful sentences. Here 5 questions were provided and it is estimated that another 5 minutes are provided for the answering of the questions with an average of 1 minute per each question. Question number 10 had four items under which the student was required to choose any two questions and the write them according to the instructions given. These were essay type questions. We assume that 60 minutes were supposed to be provided to answer the two questions, upon which 10 minutes were reserved for reading all the questions, making decision on which questions to be answered, listing down the points and finally answering them.

The last section composed of 4 questions under which the student was to give response to the readings done in the course of study. A total of two questions were supposed to be answered from this section and an assumption of 60 minutes is provided for answering these questions considering 10 minutes for reading, choosing the questions, listing down the points, and the remaining 50 minutes for the actual writing, with 25 minutes per each question. This information is given in summary form in Table 2 below.

Table 2. Item type Distribution

Sn.	Part	Item type	Number of items	Minutes
1.	Language use	Scrambled sentences	5	5
		Matching items	5	5
		Essay writing	2	60
2.	Vocabulary	Completion	4	5
		Filling gaps	8	6
		Short answers	4	6
		Correcting sentences	8	8
		Rewriting sentences	4	5
3.	Comprehension	Multiple choice	4	20
		Completion	3	5
		Summarizing	5	15
4.	Orthography	Essay	2	60
			Administration:	20
			Total	220

As it is indicated from table 2 above, 40 minutes more were supposed to be added so as to help the examinees whom English is their second language to achieve the expected goals. It seems that the examiners just provided the time for the examination generally without considering the actual time for each question. This practice had happened for a number of years that the time allocated for the Form Four National Examinations is 3 hours regardless the item difficulty levels or the number of items per subject. Ideally, as Kluitmann (2008) suggests, the time allotted for the completion of the tasks depends on the difficulty level. Khattri, Reeve, and Kane (1998) also list duration of examination among the five "dimensions" in the criteria for task specifications to be used in designing performance tests. This time allotment is also seen as crucial by Douglas (2000) and must feature in the front page examination rubric.

3.4 Giving Examination Instructions

Test instructions, as Harris (1969) suggests, should be brief, simple to understand and free from ambiguities. They should be accompanied by sufficient examples so as to be clear to the learners of all levels. Instructions on guess work should also be provided, length of time will be allowed for the test and its several parts are to be indicated. Harris (1969) further posits that the purpose of test directions is to allow all examinees to begin the problems equally. If directions to the test are poorly constructed, they contribute to poor performance of the examinees and as a result they are being blamed as fool while the problem is with the examiners.

From this paper, four general instructions were provided on the cover page, indicating sections, what to be answered from each section, the forbidden use of phones and calculators, and lastly students were directed to write examination number on every page of their booklets (but no instruction directed them where to write their answers). From each section and sub-section, different instructions were given on how the questions were to be answered. No examples were provided from any section and sub-sections.

In Section C (Language Use) the examinees were instructed to answer a total of four questions (i.e. 8 and 9 and any other two). Here the instructions were doubled. At the beginning of the section the instruction reads, *Answer question 8, 9 and select any two (2) in question 10*. In question 10 the instruction reads, *Answer two questions from the alternatives A, B, C and D*. Apparently there is no ambiguity from these instructions except for unnecessary repetition. It was probably assumed that the candidates, having done the two questions, would likely have forgotten the earlier instructions. In that case then the two first questions answered (8 & 9) were supposed to have different instruction from question 10 to avoid unnecessary repetition. It is seen from the evidence that the instructions were not clearly written and no examples were provided so as to give guide to the examinees on how to tackle the questions provided.

3.5 Reproducing the Test

It is advised that test materials be reproduced as clearly as possible so as not to affect students' performance. This was the CSEE paper, which was done country-wide and hence it

was a large scale testing. The paper was legibly readable and no single letter was misspelled. This shows that great care was taken when reproducing the examination. Multiple-choice items were provided on a single page and no item began on one page and continued on the next. On top of that, the comprehension passage that was provided has its subsequent questions adjacent to it to the extent that it was easy for the candidate to go back to the paragraph for the clarification of the specific question being asked, so as to provide the correct answer.

For the task-completion questions, no guidelines were provided on where the examinee was supposed to write his/her response. On the cover page there was an instruction for writing examination number on every page, but there was no instruction directing the candidate to write all your answers from the booklet(s) provided. This confuses the examinee, since those blanks provided on question paper may lead the student to answer the questions in them (blanks spaces) and others (such as essay questions) in the booklet. So, clear instructions on where to provide the answers are very important.

It is advised by Harris that, at the bottom of each page there should be indication on whether the examinee is to continue on the next page or stop his work until directed to do so. From the paper being discussed, no such directions were provided to the examinee.

The pages of the examination paper were glued together correctly. No staples were provided to the sheets, but the pages were put together. It is observed that stapling of the sheet is very useful since through protecting then from being detached, they prevent students from cheating as they may exchange the inner sheets so as to get answers from one another, and others, may take the inner sheets outside to be solved by others, or may ask permission as if they are going to the washroom while hiding the inner pages so as to solve outside and return with answers.

Having discussed these items, it is good now to judge upon the examiners fidelity per each section. Starting with the course objectives as indicated earlier, this examination lacks several items. The item of communicating effectively with other speakers of English both inside and outside the country is not addressed. It is believed that the possible way to measure ones' speaking ability is to hear how he/she communicates with others in English. It is true that it is not possible to measure all the items per single examination, but those basic skills are to be given priorities. Most of the Tanzanian learners of English have more problems in speaking than in writing. The leaner may know the rules of grammar in writing but won't be able to apply them when speaking. It could be very helpful if plans were put in place to start measuring listening and speaking skills. It is true that this kind of item takes a lot of time, but with good plans this can be possible.

Another point to discuss here is about the use of one's knowledge of English in furthering one's education. This also is among the objectives of the course in general. From the given paper, this item is lacking, since the paper doesn't show any clue that there is an indication which shows that the learner had gone through a certain course which will help him/her for further studies. How can the paper of terminal nature like this prove the learner's ability to communicate if there is no question to give the evidence? Were the examinees tested before

sitting for the final examination? To what extent? How were they measured? Were there any prior plans from the Ministry of Education and the Examination Council to measure the examinees before sitting for the written paper?

Another area on which the general objectives were not met is on the part of communicating effectively with other speakers of English both inside and outside the country. There is no evidence from the paper that shows this item was tested. We are supposed to see some kind of dialogue on which the examinees are to fill some parts of the written dialogue so as to measure their ability in conversation. If the learners were to communicate with speakers outside the country, where will they find them so as to communicate with? If not what is another alternative which will enable them to communicate with outsiders? Is speaking with the people outside the country a standard on to which learners should be measured?

On the part of dividing course objectives into their components, four components were measured which are comprehensions and summary, language use, pattern and vocabulary and response to reading. Though several components were measured, in general, this examination tested reading and writing skills from each component as indicated from the table provided previously.

Speaking and listening skills were not measured. Strategies should be devised on how to measure learner's speaking and listening skills. Harris provided some examples on testing different items including speaking and listening, and this is what TOEFL and IELTS examinations are doing. All the four skills are to be measured so as to judge correctly in which level does the learner is.

3.6 The Time Factor for the Examination and Speediness

As it was shown earlier, the allocated for the examination didn't consider the number of items provided to be answered. The time was taken generally that any CSEE paper is done in three hours without any consideration of the type of examination and number of items included in that specific examination. The time provided didn't even consider that the examinees do differ in their speed upon performing given tasks. Extra minutes are to be provided to those learners who are slow so as to judge them correctly on their performance.

Generally, the observed information can be summarized in Table 3 below, indicating percentage wise the fidelity of the examiner adhering to examination setting steps.

Table 3. Percentagewise Extent of the Examiner Fidelity in Examination Construction Steps

Steps	Components	Yes (in %)	No (in%)
Course Objectives	Express in spoken and written English	75	25
	Read widely for pleasure and information	100	0
	Perform variety of tasks	100	0
	Use knowledge of English to demonstrate awareness	100	0

	Use knowledge of English for further education	0	100
	Communicate effectively in English	0	100
Dividing General Objectives	Listening	0	100
	Speaking	0	100
	Reading	50	50
	Writing	75	25
Time Factor and Speediness	Time for the whole examination	75	25
	Time per item	0	100
Examinations Instructions	General instructions	75	0
	Specific instructions	75	0
	Examples	0	100
Producing and Reproducing the examination	Readable texts	100	0
	Purpose of the test	100	0
	Method of recording responses	0	100
	Method of changing answers	0	100
	Information for guess work	0	100
	Proscription not to write in test book (reusable)	0	100

4. Conclusion and Recommendations

From the given observations and explanations it is found that this examination was not at par with ideals of examination setting steps because the course objectives were not all met, the components of general objectives were partially met as speaking and listening were not tested. The time provided for the test was insufficient considering the items indicated. Some of the instructions were poorly constructed and others were not there completely. No examples were given to the examinees as guide to answer the questions provided.

It is highly recommended that serious measures should be taken when constructing the test. There are many factors which contribute to the poor performance of the examinees. In many cases examinees have been blamed for the poor performance which might not be their fault but the examining body (NECTA, in our case). Every step in examination construction should be considered as important as test itself because examination without the proper instructions is not exam at all. Harris (1969) suggests that if the test is poorly constructed, the specific skills intended to be measured will not be proved, only by chance.

All the four language skills should be measured. In most cases only reading and writing are those which are measured frequently. Ways of performing other skills should be developed so as to ensure that learners are competent in all skills. It is true that measuring speaking and listening is a little challenging when compared to writing and reading, but with plans things will work out. If it is not possible to measure these skills in the final paper, arrangement

should be done so as to make them tested during the course of study or nearing the final examination since it is time consuming.

On testing speaking and listening, language experts apart from subject teachers are to be invited so as to make the examination more valuable and reduce to the teacher testing the students knowing that he/she is measuring his/her own students. Just as outsiders come for examination supervision, the same should apply to this kind of testing.

Most of the examinations to be provided whether classroom examinations, placement examinations, selection or certification must be guided by the objectives of the course and all the procedures for producing good test should be followed. This will increase efficiency in teaching and learning and the goals will be achieved.

The examinations should be constructed in such a way that there is no single chance for students to blame examiners for constructing poor examinations, the blame should be on their own either for not following instructions properly or for not understanding, and not because of ambiguity caused by the examiners.

References

British Council (n.d.). *The IELTS Question Paper Production Process*. University of Cambridge: Cambridge English Language Assessment.

Brown, J. D. (2000). Observing Pragmatics: Testing and Data gathering Techniques. *Pragmatics Matters, Vol.1, No.* 2:5-6.

Carroll, J. B. (1968). "The psychology of Language Testing". In A. Davies (Ed.), *Language Testing Symposium: A Psycholinguistic Approach*. London: Oxford University Press, 1968. Chapter 4, pp. 46-69.

Carroll, J. B. (1973). Foreign Language Testing: Will the Persistent Problems Persist? Presentation at the ATESOL Conference, Dublin, Ireland, June 26-29, 1973.

Harris, D. P. (1969). *Testing English as a Second Language*. New York: Macmillan.

Khalifa, H., & Weir, C. (2009). *Examining Second Language Reading. Studies in Language Testing 29* Cambridge ESOL and Cambridge University Press, Cambridge.

Khattri, N., Reeve, A., & Kane, M. (1998). *Principles and Practices of Performance Assessment*. Manwah, NJ: Lawrence Earlbaum Associates.

Kluitmann, S. (2008). Testing English as a Foreign Language Two EFL-Tests used in Germany. Unpublished Doctoral Thesis (ELT), University of Albert-Ludwigs.

Milanovic, M. (2002). Common European Framework of Reference for Languages: Learning, Teaching, Assessment Language examining and Test Development Prepared under the Direction of Strasbourg (A.L.T.E.). Language Policy Division.

MOEVT. (2010). English Language Syllabus for Secondary Schools Form I – IV. Dar es Salaam: TIE.

O'Grady, W., Dobrovolsky, M., & Katamba, F. (1996). *Contemporary Linguistics. An Introduction*. Essex: Pearson Education Limited.

Ross, K. N. (2005). *Overview of Test Construction Quantitative Research Methods in Educational Planning*. UNESCO International Institute for Educational Planning.

Salisbury, K. (2005). The Edge of Expertise: Towards an Understanding of Listening Test Item Writing as Professional Practice, Unpublished Doctoral Thesis, King's College London.

Stern, H. H. (1983). *Fundamental Concepts of Language Teaching*. Toronto: Oxford University Press.

Suen, H. K., & McClellan, S. (2003). 'Item construction principles and techniques'. In N. Huang (Ed.) *Encyclopedia of Vocational and Technological Education,* Vol. 1: 777-798. Taipei: ROC Ministry of Education.

Copyright Disclaimer