

**Language Testing and its Washback Effect on Teaching:
The ESLC English Language Examination in Focus**

Dejenie Leta

Department of Foreign Languages and Literature

ABSTRACT

This study addresses the issue of the washback effect of language testing on the teaching of the language as well as on the instructional materials (to be) used. The effect of the ESLC English is examined from this perspective and a recent study to redress the alleged deficiencies of the said examination presented.

In the mentioned study, a new examination was designed comprising two parts, each of which was intended to be better geared to each of the two tasks of ESLC English: certifying the completion of secondary school English (as an achievement test), and screening deserving secondary school leavers for university education (as a proficiency test). In an attempt to satisfy each of these requirements, the first part sampled the contents, methods, and formats of the secondary school English textbooks and curriculum while the second part simulated a reasonable sample of the types of real-world academic activities that first year university students would use their English for.

Both parts of the new examination were then administered to over 1000 ESLCE candidates of the 1986/87 academic year. Furthermore, the secondary school English scores of some 300 of these candidates as well as the views of a representative sample of people concerned were gathered.

Both the empirical and judgemental investigations revealed that the first part of the new examination was better than the currently used one as both an achievement and a proficiency test, whereas the second part was better at predicting university performance. Moreover, a combination of the candidates' secondary school English scores and any of the two parts of the new examination was found to be better in this aspect than any of the individual score used in the study.

Relevant conclusions were drawn and implications for future undertakings discussed.

Introduction

This paper has three major aims. Firstly, it attempts to highlight the importance of (language) testing and its washback effect on both teaching and the preparation and/or revision of teaching materials. It then tries to survey the alleged major weaknesses of the ESLC English language examination and the possible sources for these weaknesses. Finally, it reports on a recent research conducted to redress the deficiencies of the current ESLC English language examination and its format by attempting to introduce feasible innovations into possible future testing of the language in the light of recent developments in the field.

[Important: The mentioned "recent" research was conducted about five years ago and, so, should be understood in that context.]

I. Testing and its Washback Effect on Teaching as well as on the Designing or Revision of Teaching Materials

Developments in language testing cannot be viewed in isolation as teaching and testing are two inseparable components of the entire teaching/learning process. That there is such a very close relationship between the two is explained by Heaton (1982:5) as follows:

Both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other.

Davies (1968:5) is of the opinion that "a good test is an obedient servant since it follows and apes the teaching." According to Hughes (1989:2), on the other hand,

We cannot expect testing only to follow teaching. What we should demand of it, however, is that it should be supportive of good teaching and, where necessary, exert a corrective influence on bad teaching,

thus attesting to the fact that teaching and testing go hand in hand influencing each other.

From the practical point of view, Alderson (1986) discusses how the close relationship between teaching and testing has been exploited to the extent that tests have been used as instruments of innovation. He elaborates on this by citing three examples: the Royal Society of Arts' (RSA) Communicative Use of English as a Foreign Language (CUEFL) as an example of a test resulting from the introduction of notional-functional syllabuses in language teaching; the teacher-led Graded Objectives in Modern Languages

scheme showing how tests motivate both teachers and students, as well as the Sri Lankan National Certificate of English which led not only to a closer working relationship between teachers and testers but also to changes in teaching.

It is this effect of testing on teaching and learning that is known as the "washback" effect of testing (Hughes, 1989:1). Washback can be positive or negative depending upon the quality of the test or examination used. According to Pearson (1988), the objective, multiple-choice dominated ESLC English is an example of tests having a negative washback effect on teaching.

In contrast, Pearson (1988) also refers to other situations where tests have been instrumental in effecting improvements in teaching through their positive washback effect. A good example in this regard is that of the Sri Lankan Certificate in English which, despite numerous constraints that are quite similar to ours, introduced innovations to its testing and this, Pearson hopes, may have eventually contributed to the improvement of teaching.

The other instance that he (Pearson, 1988) mentions is that of the Scaling Test of the Junior Secondary Education Assessment system introduced to Hong Kong in 1982 which, according to him,

led to syllabus revision, the design of new textbooks with different goals and objectives from those currently in use, and, as a consequence, to new class-room methods and techniques which stress the use of authentic English in purposeful and interactive way (Pearson, 1988:100).

In Hong Kong, the importance attached to tests "as a direct determinant of class-room teaching methods and techniques and the design of instructional materials" made Johnson and Kin-Ling (1981:285), the designers of the aforementioned test, feel confident that the adoption of the new test will be "a precursor of syllabus revision at all levels of the school system." Ibid. This, they believed, will eventually:

lead to the establishment of a new equilibrium in the interdependence of language teaching, language testing, and instructional materials which exploits the opportunities for learning, teaching, practice and use of English for communication purposes in the particular setting (Ibid.).

Mention can also be made of several other examples or situations where the designing and putting into effect of new English language examinations have given rise to an improved teaching practice and to a parallel improvement in syllabus design and teachers training.

Attempts made by Hughes (1988b) to introduce a new pragmatic English language examination to Bogaziçi University (Turkey), and that by Henning et al (1981) to Cairo University are other examples of such undertakings aimed at introducing similar revolution to the English language teaching and testing practices of the said countries.

II. Major Weaknesses of the ESLC English Language Examination

What have been enumerated above attempt to highlight the important role that testing plays in improving both teaching and teaching materials. Unfortunately, though, the ESLC English does not seem to measure up to these expectations. In fact, it appears that there is a strong feeling, especially among educators, that it has not only failed to be an effective measure of candidates' mastery of the language, but has also negatively influenced the teaching and testing of English in Ethiopian (secondary) schools.

Expressed succinctly, one of the main criticisms against the English language examination currently in use is that it does not adequately accomplish the two major tasks that it is supposed to: certifying the completion of secondary school English (as an ACHIEVEMENT TEST), and serving as a selection criterion for university admission (as a PROFICIENCY TEST).

The other major criticism, as stated above (Pearson, 1988), is that its purely multiple-choice format is exerting a negative washback effect on the teaching and testing of English in Ethiopian secondary schools. There is also a feeling that this adverse effect is having its impact on the teaching/learning and testing processes at the junior and elementary school levels as well.

If the ESLC English is fraught with all these deficiencies, then, it is difficult to assume that it is satisfying its minimum requirement as a valid yardstick for the intended purpose(s), let alone being instrumental in introducing the desirable changes and/or improvements to the teaching/learning process.

As any undertaking to redress any of these deficiencies should spring from a full understanding of what the likely causes could be, an attempt will be made to outline hereunder the major factors that could contribute to the alleged weaknesses of the ESLC English.

A. The Lukewarm Attention Paid to Testing and the Need for a Quality Controlling English Language Committee

The evidences enumerated under Section I above attest to the fact that testing is a very strong force or agent for desirable changes in teaching. Even then, what is equally important is the realization that testing can effectively execute such a formidable responsibility if and only if it can satisfy minimum quality standards.

One of the important steps towards ensuring such quality standards in tests is the paying of due attention to testing itself during the training of teachers. Although, as explained above, testing is as important as, and by no means inferior to, teaching, yet, our language teachers are not given any training in (language) testing in TTT's, in inservice and/or extension programmes, or in seminars. It is very surprising that even at the undergraduate and post-graduate levels, the situation is not any better as, for instance, the one time three-credit course on language testing has ceased to exist as a course and made to just form an insignificant component of the language teaching methodology course.

This means that no condition has been set in our educational system whereby teachers are adequately equipped with the techniques of testing or evaluation even though these form a very essential component of the entire teaching/learning process.

Neither is there any systematic way of checking and/or assessing whether or not our tests, especially those at the national level, are the appropriate yardsticks for adequately measuring what we would like them to. A rigorous assessment of the quality of our tests/examinations would necessitate the creation of an English Language Committee within the ESLCE Board (possibly comprising members representing the Department of Foreign Languages and Literature, language testing expert(s), secondary school English language teachers, as well as textbook writer(s)). Such a committee, if formed, would be responsible for the necessary monitoring and quality control required all through the designing, piloting, administering and validating stages of the examination. By so doing, it would be able to build reliability and validity into the examination right from the very beginning and to also keep a constant check on its quality for the necessary improvements as deemed appropriate. It could also help in maintaining standards between the various years' administration of the examination, an aspect which has been non-existent so far and so casts doubt on whether the levels of performance represented by A's, B's, C's, etc. during a particular year are comparable to those of other years.

B. Other Problems Emanating From The Nature of Tests and the Linguistic Theory That the ESLC English is Based On

If we look at the sources of the weaknesses of the ESLC English language examination from a different angle, we find that there are other contributing factors which spring from testing in general and from the theoretical linguistic assumptions on which the ESLC English language teaching and testing seem to be based.

To start with, just like any test, it is a very minute representation or sample of what has been taught. Assuming that the examination is based strictly on the assigned English language textbooks and syllabus, the 90-minute examination of some 120 items is expected to assess what has been taught for 12 years. What makes this problem even worse, especially in the case of the ESLC English, is the fact that there are no checking mechanisms (as given under II-A above) that could guarantee that the contents and methods of the assigned textbooks and syllabus are adequately sampled, or that the selected items possess the appropriate quality to assess those contents and methods.

Furthermore, the failure of the items to be adequate representatives of the contents and methods covered in class also raises the serious questions of generalizability. When it is very doubtful as to whether or not the items of the examination are a fair representation of the English language covered in (secondary) schools, the situation casts more doubt over the generalizability of the ESLC English grades or scores as true reflection of candidates' (secondary school) English mastery.

Moreover, the situation is further complicated if, as has been the assumption to date, the ESLC English is considered as a proficiency test which can serve as a screening device for university/tertiary level admission. As a forward-looking test, any proficiency test worthy of the name is expected to adequately assess how candidates can put their English language mastery to effective use in the real-world or realistic tasks pertaining to the setting - be it university or otherwise - that they aspire to join. From this perspective and given what the current ESLC English is, it is hardly possible to believe that either its content or the method/format measures up to the minimum standards of a proficiency test/examination. The generalizability of its scores/grades in this regard thus becomes almost nil.

In fact, this could lead to the conclusion that it would be erroneous to expect the ESLC English in its present form to be a measure of candidates' English language proficiency since, as Carroll (1980:8) puts it, tests aimed at measuring the ability to use the language - as opposed to knowing the proper usage - should not be based on the "selection of items chosen on linguistic grounds alone."

Secondly, the ESLC English seems to inherit its weakness partly from the linguistic theory/theories it is founded on. Davies (1982) gives a number of drawbacks of the pre-scientific and structuralist periods of language testing, both of which are considered out-dated in modern tests, but, on both of which it (the ESLC English) appears to be based. Two of these drawbacks are quite relevant here:

- (1) Both (that is, the pre-scientific and the structuralist approaches) focus on bits of language that are easy to devise and assess; i.e., emphasis is on the discrete-point elements rather than on integrated language; and
- (2) they deal with de-contextualized skills rather than a systematic consideration of purpose and strategy - an erroneous approach that delayed the method of teaching and testing of language skills in a manner that can approximate natural ways.

Thus, the current ESLC English model and techniques reflect out-dated views regarding language teaching and testing.

Despite its advantages of being simple to administer and easy to mark, the structuralist approach to language analysis for testing purposes - which is more dominant in the case of the ESLC English - has a number of drawbacks. To start with, candidates are involved in little or no active production. Moreover, it is not easy to know if a score is a reflection of a candidate's actual performance on the test or is influenced by the method effect. Furthermore, it is difficult to know if the items 'work' as items (Baker, 1989:37-8)

One outcome of the structuralist approach to language teaching and testing is the excessive use of the multiple-choice format. In fact, this format dominates all classroom teaching, exercises, and tests to the extent that the ESLC English language examination and its format gain importance from year to year until, at the final year of secondary education, they almost assume the role of a teaching material. It thus seems that it is the strict adherence of the ESLC English to the solely multiple-choice format that has induced candidates and, possibly, even their English language teachers - despite their continued resistance (Forrest, 1975:244-5; Ministry of Education, 1983)-, to feel no need for practising the language in meaningful communication in class.

Hughes (1989) argues that the multiple-choice format tests only the recognition of knowledge which is in part a question of construct validity. He also believes that it encourages guessing and/or cheating. Furthermore, it severely restricts what can be tested and it is often very difficult to write successful items.

Hughes adds that,

where a test which is important to students is multiple-choice in nature, there is a danger that practice for the test will have a harmful effect on learning and teaching (Hughes, 1989:61-2).

His general advice is, therefore, to avoid the excessive and/or indiscriminate use of the potentially harmful multiple-choice format.

C. Studies Conducted on the ESLCE and/or the ESLC English

Most of the studies conducted regarding the ESLCE, in general, and/or the ESLC English, in particular, focus mainly on its capacity as a proficiency test to investigate whether or not it adequately predicts success of university.

Quite a large number of these studies have revealed that the ESLCE has little or no potential as a predictor of university success (Kortaw, 1964; Kehoe, 1964; Langmuir and Bowers, 1967, for instance). Still others showed a low to average correlation (Tracy, 1965; Bowers, 1969, for instance), and the often cited study in this regard - because of its relatively higher correlation figures - is that of King and King (1970) where it was found that:

Total ESLCE scores in an average of five subjects shows the highest correlation with GPA ($r = .452$). The second highest correlation of test scores with GPA is for the ESLC English test ($r = .420$) (King and King, 1970:50).

As an outcome of these studies, the weaknesses of the ESLC English as a potential predictor of university performance are attributed to the examination being made to serve as both an achievement test and a proficiency test (Javis, 1969; Madsen, 1976; Tesfaye Gashaw, 1982; Ministry of Education, 1986).

According to the recommendations of the Evaluative Research of the General Education system in Ethiopia (Ministry of Education, 1986) which, relatively speaking, is more exhaustive and recent,

1. The designing of the ESLCE as an achievement test should be the responsibility of the Ministry of Education, with its weight accounting for 50 per cent of candidates' overall results, while the remaining 50 per cent should be allotted to their four-year secondary school scores in the particular subject.

2. The recruitment of students for tertiary level education must be based on an entrance examination to be prepared by the institution(s) concerned as well as on the results of ESLCE achievement test and class tests.
3. Courses on Measurement and Evaluation should be incorporated into all pre-service teacher training programmes (Ministry of Education, 1986:3)

III The Present Research

A. Introduction

An attempt has been made so far to highlight the crucial role that (language) testing plays in the teaching/learning process as well as some of the observed weaknesses and/or problems associated with ESLC English in meeting that standard.

Mention can also be made of other problems that aggravate the situation in Ethiopian schools. The major ones, as mentioned in several studies in the past, include the large class size that the English language teacher is made to confront, the shift system, the quality, variety, and quantity of the teaching materials in use, as well as the quality and morale of the teachers. It is obvious that each one of these problems would have a great impact on the teaching and testing of English in schools.

The solutions to these problems demand an exorbitant amount of money which the present economic condition of the country may not be able to provide at the moment. Even then, the present-day teaching and testing of English in Ethiopian schools appear to be in such a critical condition that we cannot afford to wait for a desirable economic condition that could warrant corrective measures in all spheres concerned. Therefore, possible changes and effective measures have to spring from within the teaching and testing world, and the research discussed hereunder was a modest attempt to indicate what shape and form these changes could take.

B. Aims of the Study

The major aim of the present study was to look for ways of improving the English language component of the ESLCE so that it would be a better measure of achievement and proficiency and, by so being, exert parallel improvements on the classroom teaching of the language at the secondary school level.

As the study was exploratory in nature, it was felt necessary to:

1. design a two-part examination with each part adequately sampling each of the real-life situations that ESLC English is expected to mirror: the classroom teaching and testing of English in Ethiopian secondary schools (as an ACHIEVEMENT test), and the use of English for the various academic tasks at the university level (as a PROFICIENCY test);
2. sample widely not only the contents but also the methods/formats and activities in which English could be employed;
3. assess the effectiveness of candidates' secondary school English language scores as supplementary measuring instruments;
4. do an opinion survey on what the people concerned feel about the new examination and the already existing one, and
5. analyse both the qualitative and quantitative data in order to check whether or not the new measuring instrument (namely, the new examination) is better than the existing one in accomplishing what an ESLC English language examination is ideally expected to.

The assumption made in the designing of the new examination as well as the introduction of a balance in the testing methods to closely relate to the purposes of the test was that a test which attempts to approximate the real-life purposes and activities that English is used for could exert a positive washback effect on the classroom teaching and testing of the language.

C. Guiding Principles

The main guiding principles in the designing and validation of the new examination included:

1. **Washback:** Any test/examination hoped to exert a positive washback effect on teaching should, according to Hughes (1989: 44-7), not only test the abilities whose development one wishes to encourage but also sample widely and unpredictably. Furthermore, an achievement test should be based on the objectives for teaching the language.

These principles were adhered to in as much as possible while designing the new examination.

2. How Best to Tap Language Mastery: An exploratory research like the one under discussion here should employ all methods and techniques possible in the designing of a test so as to come up with what could prove to be the best for the purpose. In this connection, Davies (1985: 180) believes that such a test should take three main forms:

- (a) The Structural Approach - forming the foundation for the discrete-point tests,
- (b) The Functional Approach - embodied in the communicative competence model, the notional-functional curriculum, and the need to teach and test pragmatism, and
- (c) General Proficiency - based on the belief that "individuals vary in possessing immeasurable amounts of an indivisible body of knowledge" (Ibid). It argues in favour of a general factor underlying batteries of tests.

Spolsky (1985) also takes a similar stand. Both Davies (1985) and Spolsky (1985) believe that anybody who knows a second language may be assumed to have all of the three kinds of knowledge as none of them can be claimed to be completely adequate in itself. This means that any test worthy of the name should try to exploit all the three approaches for an investigation into the best methods and/or combination of skills that could help in effectively measuring candidates' proficiency in English.

Accordingly, it was planned that the achievement part of the new examination should represent the Structural Mode, the Cloze and Dictation subtests should cover the Functional Mode, and the proficiency part should embrace the General Proficiency Mode.

As regards the formats or techniques of testing to be employed, it has already been said that the multiple-choice format is the sole method adopted by ESLC English and that, as a result, this is exerting a negative washback effect on the teaching and testing of English in Ethiopian schools. It is moreover believed that such a total or excessive reliance on the multiple-choice format "creates a passivity syndrome and yields only marginal returns" (Fischer, 1982: 201) and that it also involves developing 'test wiseness' - the capacity to score more marks from a test by responding to cues (Wood, 1987:55).

To avoid or minimize such an adverse effect of the multiple-choice device, modern language testing advocates the use of a balanced application of all methods possible (Fischer, 1982; Hughes, 1989; Wood, 1987) so as to ease off what is known as the "method effect" that could have the potential of distorting candidates' performances and/or scores in an examination.

3. **Emphasis on Validity:** According to Carroll (1980: 1),

...the ultimate criterion of language mastery is the learner's effectiveness in communication for the setting he finds himself in.

This means that any test that is expected to assess this mastery must be, above all, valid for the particular purpose and the particular group in focus (Gay, 1980: 165).

Unfortunately, however, the ESLC English aims at enhancing reliability at the expense of validity, and this has been one of its stumbling blocks in its attempt to serve at least as an achievement test, let alone its potential as a proficiency test. The present research was, therefore, aimed at reversing this order and giving validity its rightful place while at the same time trying to minimize or avoid any sacrifice in reliability. In this connection, Weir (1988:35) gives the following piece of advice as regards the rivalry between validity and reliability:

It is sometimes essential to sacrifice a degree of reliability in order to enhance validity. If, however, validity is lost to increase reliability, we finish up with a test which is a reliable measure of something other than what we wish to measure. The two concepts are, in certain circumstances, mutually exclusive, but if a choice has to be made, validity "after all, is the more important."

This should, therefore, make one of the major distinctions between the current ESLC English (which appears to favour reliability to validity) and the new examination, especially the proficiency part, which purports to assess candidates' potential in using their English language mastery for real-world purposes (of the institution they wish to join), namely, tasks related to university academic pursuit. It was hoped that the new examination thus designed would be more fit to meet the required face, content, construct, and predictive validity demands than would the current ESLC English.

4. Practicality and Appropriacy. Because the objective nature of the ESLC English examination pays attention to the "form" of the language rather than to its "function", the examination has so far made no attempt to include all or any of the important skills such as writing, listening, and speaking. It is very likely that this has been the case simply because, as Hughes (1989:23) correctly observes, "too often the content of the tests is determined by what is easy to test rather than what is important to test." [Emphasis mine].

It is one thing to abide by the dictates of practicality such as economy, scorability, and administrability (Finocchiaro and Sako, 1983: 30), but quite another to reject outrightly the use of any device(s) that may appear to be both appropriate and feasible. One thing that needs to be borne in mind is the fact that time is a crucial factor in the domination of a particular teaching and/or testing method, and the past twenty or so years that the ESLC English format has been made to cherish appear to be far too long. It is thus high time that we realized that the ESLC English sole multiple-choice format has outlived its time and that we sought for ways of improving or modifying it.

In this regard, Underhill's (1981: 23) remarks appear to be apt:

...what kind of test you use should be determined pragmatically by the purposes for which the test is to be used, the resources you have available for construction, administration and marking, and what you intuitively feel will have the highest face/content validity for testers and testees alike.

Such an approach to the designing and use of a test is hoped to strike a balance between the two extreme viewpoints that Davies (1985) terms as "excessive conservatism" and "unthinking radicalism" that are likely to tempt (language) test designers. The new examination was planned to be based on this principle of a balanced approach.

5. Triangulation: Being exploratory in nature, the study had to employ the method of triangulation. This necessitated, on the one hand, the gathering of the views of those concerned such as candidates, their secondary school English teachers, textbook writers, and university freshman instructors regarding the current and the new examinations, and, on the other, subjecting all the instruments of the research to all possible judgemental and empirical investigations.

On the whole, the designing of the two parts of the new examination was based on the aforementioned guiding principles. As a result,

- a) the achievement part of the new examination was to be different from the ESLC English currently in use in its:
 - (i) attempt to strictly mirroring the contents and formats of the secondary school English textbooks and curriculum;
 - (ii) distribution of items /subtests/ subskills in proportion to the weighting or importance each is accorded in the sources mentioned [(i) above], and
 - (iii) use of not only the multiple-choice format but also other methods such as true/false and short-answer forms, thus providing candidates with varied chances to demonstrate the actual magnitude of their mastery of the language.

- b) The proficiency part of the new examination was to be different from the current ESLC English in two distinct ways:
 - (i) it (the former) attempted to look forward - to Addis Ababa University - in investigating how successful secondary school leavers could use the English they have hopefully mastered at the secondary school level for meaningful academic tasks required at the university level such as lecture note-taking, taking notes from materials read, and editing lecture notes, and
 - (ii) the candidates were expected to write out their answers rather than blacken or fill in corresponding circles as practised in ESLC English.

This was based on the hope that the contents, formats, and methods of the proficiency part of the new examination could shed some light on the types of changes that could eventually be introduced to the teaching and classroom testing of English at the secondary school level, thus encouraging the teaching and testing of English in the right context: COMMUNICATION.

D. Research Methodology

1. Instruments Used

(a) **Examinations:** Three examinations were used for comparative and validation purposes: the 1986/87 ESLC English language examination ("ESE" hereafter), the achievement part of the new examination ("NE1", hereafter), and the proficiency part of the same new examination ("NE2", hereafter).

At the initial stage, NE1 comprised 162 items broken down into Reading Comprehension, Cloze, Vocabulary, Sentence Comprehension, Grammar, Spelling, Mechanics, and Controlled Writing, to be tested in Multiple-Choice, True/False, Gap-filling, and 'write-short-sentence' formats.

110 of the items were prepiloted on a heterogeneous group of 22 individuals at Lancaster University, UK, comprising native and non-native speakers of English. The hope was that such a try out would help in obtaining introspections on the general impact of the test and in identifying items which are palpably unsatisfactory (Davies, 1990: 12).

Accordingly, based on the performance and comments of the participants, 20 of the items were rejected and the length of the comprehension passage almost halved. Moreover, the responses of the native speakers and the top scorers served as model answers, especially where the responses to certain items proved to be contentious. At the end of the prepiloting, therefore, NE1 had a total of 142 items to be administered to the target candidates.

NE2, on the other hand, was broken down into Reading Comprehension, Writing, Listening, and the two indirect tests of proficiency, namely, Cloze and Dictation. On the whole, it was worth 233 points, and care was taken to make the marking as reliable as possible by making the would-be answers objective (of the one-word or short-answer types).

Although NE2 was not subjected to any prepiloting, comments from colleagues helped a great deal in making changes and/or modifications wherever required.

b) **Questionnaires:** Questionnaires were designed for the candidates (in Amharic), their secondary school English teachers, as well as for university Freshman English and content instructors - all aimed at getting the views of the sample respondents on the level of effectiveness of ESE, NE1 and NE2 as both achievement and proficiency tests. All these questionnaires were prepiloted before they were put to actual use.

c) **Interview:** The director of the ESLCE office and a member of the English language textbook writers were interviewed, the former to detail the duties and responsibilities of the ESLCE Office/Board, and the latter to compare the contents and methods of secondary school English language textbooks and syllabus with those of ESE.

d) **Academic Records:** The four-year secondary school English average scores ("SSAVENG", hereafter) as well as the 12th grade English scores ("SSFYENG", hereafter) of some 300 of the candidates were collected for both concurrent and predictive validity investigations.

e) **Candidates' University First Year (First Semester) Grades:** Successful candidates' first semester Freshman English grades ("FLEN", hereafter) and the first semester overall GPA for all (five) freshman courses ("UNIVCUM", hereafter) were also collected to serve as criteria to assess the predictive potential of ESE, NE1, NE2, SSFYENG, SSAVENG, or of any meaningful combinations of these variables.

2. Samples in the Study

(a) **Schools:** As administrative constraints did not allow the inclusion in the study of all secondary schools either throughout the country or in Addis Ababa, only 10 secondary schools in Addis Ababa were selected using the stratified sampling method.

(b) **Candidates:** Over 1200 candidates from these 10 secondary schools were randomly selected for the study.

(c) **Secondary School English Teachers:** A total of 60 twelfth grade English language teachers were involved in filling in the questionnaire designed for the purpose.

(d) **University Freshman Instructors:** 23 Freshman English as well as 12 content instructors also participated in the study.

3. Statistical Methods Employed

The major statistical methods employed include:

- Descriptive statistics (measures of: central tendency, variation),
- Correlations,
- Item analysis,
- Reliability (Internal consistency),
- Validity (face, content, construct, concurrent, and predictive),

- Factor analyses (the traditional Principal Component analysis, and factors extracted until all factors with an eigenvalue of at least 0.50 are extracted), and
- Regressional Analyses (the Step-Wise Regression, the traditional Ordinary Least Square procedure, and the Ordinal Regression Model).

Use was made of the SPSSX statistical package or computer programme for almost all statistical investigations cited above except for the item analyses which required the employment of both the SPSSX and the Fortran programmes.

E. The Findings of the Study

1. **Judgemental Assessment:** Candidates who filled in questionnaire 'A' (designed to assess ESE in isolation) expressed the view that ESE is an adequate achievement test. However, those who filled in questionnaire 'B' (designed to compare ESE and NE1) preferred NE1 to ESE in its adequately mirroring the contents and methods of secondary school English textbooks and curriculum. Likewise, the secondary school English teachers felt that NE1 is better than ESE by all counts. The only reservation they had was in the inclusion of the Cloze subtest.

University Freshman English and content instructors who assessed ESE and NE2 in the light of realistic academic tasks that candidates are expected to use their English proficiency for at the university level expressed the view that it was only NE2 that meets the requirements.

One can thus safely conclude that ESE is inferior to NE1 and NE2 both in face and content validity requirements in accomplishing its tasks as either an achievement test or a proficiency test.

2. Empirical Assessment

Table 1: Percentage Score Distribution of ESE, NE1 and NE2

Score Range	No. of Candidates		
	ESE	NE1	NE2
91% - 100%	---	---	1
81% - 90%	1	13	11
71% - 80%	21	70	32
61% - 70%	33	107	82
51% - 60%	89	197	101
41% - 50%	188	287	210
31% - 40%	376	236	353
21% - 30%	505	225	183
11% - 20%	57	44	30
0% - 10%	---	2	90
TOTAL	1270	1181	1093
SKEWNESS	1.22	.40	.17

ESE = 1986/87 ESLC English

NE1 = (Achievement) Part I of new exam

NE2 = (Proficiency) Part II of new exam

As given in Table 1 which shows the percentage score distribution and the magnitude of the skewness of the distribution of the three tests under consideration, we find that ESE does not sufficiently spread the candidates through all the ability ranges and so is more skewed than either NE1 or NE2, while NE2, on the other hand, is better than both ESE and NE1 in this aspect.

Table 2: Descriptive Statistics of ESE, NE1, and NE2

Test	No. of Items	Mean %	SD %	Reliability (alpha)
ESE	119	34.4	12.2	.89
NE1	142	42.8	15.9	.95
NE2	233	38.3	16.9	.97

[Note: A t-test showed that there was a significant difference between the means of ESE and NE1 (t-value = 38.84) and between those of ESE and NE2 (t-value = 41.62), both at 1% significance level.]

As shown in Table 2, the candidates' overall performance, as demonstrated by the percentage Mean, is higher on NE1 than on either ESE or NE2. Moreover, both NE1 and NE2 demonstrate high reliability figures (0.95 and 0.97, respectively) whereas ESE fails to meet at least the minimum reliability requirement of 0.90 that any solely multiple-choice examination -like the ESLC English - is expected to satisfy (Henning, 1987).

Table 3: Intercorrelations of ESE subtests

SUBTEST	1	2	3	4	5	6
1.Sentence Comprehension						
2.Reading Comprehension	.44					
3.Vocabulary	.46	.47				
4.Grammar	.58	.48	.54			
5.Mechanics	.27	.18	.17	.34		
6.Spelling	.09	.05	.10	.14	.06	
TOTAL	.70	.64	.70	.95	.40	.21
CORRECTED TOTAL (TOTAL-SUBTEST)	.62	.54	.58	.69	.33	.14

Table 4: Intercorrelations of NE1 subtests

SUBTEST	1	2	3	4	5	6	7	8
1. Reading Comprehension								
2. Cloze	.49							
3. Vocabulary	.57	.54						
4. Sentence Comprehension	.41	.35	.63					
5. Grammar	.55	.49	.72	.69				
6. Spelling	.21	.18	.31	.27	.38			
7. Mechanics	.28	.27	.40	.36	.48	.27		
8. Writing (controlled)	.48	.51	.67	.55	.70	.35	.40	
TOTAL	.69	.64	.86	.74	.91	.42	.53	.86
CORRECTED TOTAL (TOTAL-SUBTEST)	.60	.57	.79	.68	.82	.37	.48	.75

Table 5: Intercorrelations of NE2 subtests

SUBTEST	1	2	3	4	5
1. Reading Comprehension					
2. Cloze	.59				
3. Writing	.46	.37			
4. Listening	.38	.35	.29		
5. Dictation	.27	.19	.24	.40	
TOTAL	.61	.46	.75	.53	.63
CORRECTED TOTAL (TOTAL-SUBTEST)	.50	.42	.32	.46	.27

Tables 3, 4 and 5 show the internal validity (intercorrelation among the subtests) of each of the three tests under consideration.

According to Table 3, it is the Grammar subtest which shows the highest correlation with the Total ($r = .95$), which means that the subtest (Grammar) explains 90% of the variance in ESE. This might not be surprising given the fact that the weighting that the subtest has been accorded in ESE is about 60% (i.e. 70 out of a total of 120 items).

In the case of NE1 (Table 4), the Grammar, Writing (controlled), and Vocabulary subtests account for a high variance whereas, just like in ESE, Spelling and Mechanics prove to be the poorest.

As regards NE2 (Table 5), the Writing, Dictation, and Reading Comprehension subtests show higher correlations with the Total while Cloze is relatively the poorest in this aspect.

Based on these findings, one can see that NE1 is better than ESE in being less skewed, and in sufficiently spreading the candidates through all the ability ranges while at the sometime demonstrating better homogeneity and stability. These qualities could thus serve as good indications of the fact that NE1 is better than ESE at being a better reflection of secondary school English and in being better suited to the candidates' level. This view was further strengthened when, after item analyses, the revised version of NE1 was assessed from the point of view of the aforementioned criteria.

As regards NE2, despite its being targeted for the top few high achievers and, despite the new format that it employed to measure candidates' proficiency, it proved to be better than ESE in its spread of scores, reliability figure ($\alpha = 0.97$), and the intercorrelation among its subtests.

Other Statistics: As concerns the validity of the three tests, we have already said that ESE was found to be less face and content valid than either of the two parts of the new examination.

An assessment of the concurrent validity of the three tests showed that there is a meaningful relation between the two achievement tests - ESE and NE1-: ($r = .49$). However, there was no significant relation between the secondary school English scores and any of the three tests.

As regards predictive validity, successful candidates' first semester Freshman English grades (FLEN), and their overall cumulative GPA - for five subjects -

(UNIVCUM), were used as criteria to assess the capacity of ESE, NE1, NE2, SSFYENG, SSAVENG, or any sensible combination of these variables as potential predictors.

What seems worth mentioning at this point is the fact that, like other predictive studies, this research also had to succumb to the effects of truncation in its predictive assessment as, unlike its other assessments that considered over 1000 candidates, the predictive investigation had to be based on only 83 candidates whose freshman grades were available.

Table 6: Correlations Between the Predictors and the Criteria using ESE and the original and Revised Versions of NE1 and NE2

PREDICTOR (s)	CRITERION	
	FLEN	UNIVCUM
ESE	.37	.19
NE1 (Original)	.55	.32
(revised)	.48	.26
NE2 (original)	.51	.46
SSFYENG	.36	.14
SSAVENG	.40	.10
SSFYENG + NE1(original)	.42	.12
(revised)	.38	.10
SSFYENG + NE2(original)	.45	.10
(revised)	.40	.06
SSAVENG + NE1(original)	.55	.43
(revised)	.56	.43
SSAVENG + NE2(original)	.57	.42
(revised)	.57	.42

FLEN = Freshman English (First Semester) Score.

UNIVCUM = University (First Semester) Cumulative for 5 Freshman Courses

SSFYENG = Candidates' Secondary School 12th Grade English Scores.

SSAVENG = Candidates' Secondary School English Average (four- year) Scores.

According to Table 6 above, as regards single variables, NE1, NE2 and SSAVENG are relatively the best predictors of FLEN ($r = .55, .51$ and $.40$, respectively) while NE2 and NE1 are relatively the best predictors of UNIVCUM ($r = .46$ and $.32$, respectively). However, if compound variables are also taken into consideration for the same purpose, we find that SSAVENG + NE2, SSAVENG + NE1, NE1 and NE2 show relatively the highest predictive power of FLEN ($r = .57, .56, .55$ and $.51$, respectively) whereas NE2, SSAVENG + NE1, and SSAVENG + NE2 are relatively the best predictors of UNIVCUM ($r = .46, .43$, and $.42$, respectively), thus showing the importance of the secondary school English average scores (SAVENG) in boosting the potential of NE1 (revised) in predicting both criteria, i.e FLEN and UNIVCUM, and that of NE2 in predicting FLEN.

On the whole, though, the only single predictor of Freshman English was found to be NE1 (the new achievement test) while that of University cumulative GPA was NE2 (the new proficiency test), and ESE was relatively poorer in both cases. This fact was confirmed by means of regression analyses, especially the Ordinal Regression method.

This meant that NE1 explains 27% ($r = .55$) of the variance in Freshman English while NE2 explains 21% ($r = .46$) of the variance in the candidates' university academic performance. The strength or weakness of these relationships should, however, be gauged in the light of the language proficiency vis-a-vis academic achievement controversy.

Of the studies conducted in the past to establish whether or not there is any relationship between English language proficiency and academic achievement - using the internationally renowned English proficiency examinations such as the Test of English As a Foreign Language (TOEFL), the Michigan Test of English Language Proficiency (MTELP), and the English Language Testing Service (ELTS) -, some support while some others negate the possibility of the existence of any such relationship between the two, while still others are inconclusive (Graham, 1987:505-521).

A relatively recent study conducted in Britain showed that candidates' performance on ELTS accounts for only 9% ($r = .3$) of the variance in their academic achievement (Davies, 1984:85).

Compared with all these, then, the new English language proficiency test has demonstrated a stronger predictive capacity of the candidates' university academic achievement.

A factor analysis of the three tests showed that the two achievement tests (ESE and NE1) are unifactorial in structure while NE2 is made up of two underlying factors which may be explained as a general factor and a listening factor.

Furthermore, the various validities cited above as well as the fact that each of the two parts of the new examination is adequately assessing what it is supposed to can be proofs for the construct validity of both the achievement and the proficiency tests (of the new examination).

F. Conclusion and Implication

Based on these findings, one can accept the two main assumptions of the present research, namely, that:

- (a) Each of the two parts of the new examination is better than the current ESLC English as an achievement and /or proficiency test, and
- (b) A combination of candidates' secondary school English language scores and their performance on the two parts of the new examination is a better predictor of their university performance than is the current ESLC English.

Admittedly, as an exploratory research, the present study could show certain weaknesses. Moreover, the new examination, although innovative from the point of view of the ESLC English, could look conservative in comparison to the new testing methods employed in areas that can meet the necessary financial and skilled man-power requirements. It should, however, be borne in mind that this new examination cannot avoid being relatively conservative at this initial stage because of the various difficulties and shortcomings emanating from the contents, approaches, and method of the textbooks and syllabi currently in use, the lack of trained markers, and other matters relating to financial constraints, all of which have a decisive role in any innovative undertaking. Nevertheless, despite the impact that such deficiencies could have on any innovative undertakings, the new examination appears to fall in line with the three main recommendations of the very few recent studies advocating that:

1. the responsibility of designing and administering the ESLC English (as an achievement test) to assess what candidates have learnt at school be left to the Ministry of Education,
2. the designing and administering of an English language proficiency test/examination as a screening device to select deserving candidates for tertiary level education should be the responsibility of the receiving institutions concerned, and that

3. secondary school English language scores should be part and parcel of the criteria for assessing candidates' performance in the English language examination.

The hope is that, if and when put into effect, the innovations built into the new examination- especially to the proficiency part - would give rise to potential relative improvements in the teaching and testing of English in schools and thus lay a good foundation for desirable improvements in the teaching/learning process as a whole.

Furthermore, these improvements could have implications for English courses offered at the university level too. The teaching and testing of Freshman English, Sophomore English, and English teaching methodology courses, in particular, will have to accommodate such improvements so that a sequential ordering of the desired innovations could be put into effect.

Of course, as mentioned earlier, the need for the establishment of an English Language Committee within the ESLCE Board entrusted with the responsibility of regularly accomplishing everything necessary for quality control cannot be overemphasized.

BIBLIOGRAPHY

- Alderson, J.C. 1986. 'Innovations in Language Testing' in Portal, M. (ed), Innovations in Language Testing. Windsor, Berkshire: The NFER-NELSU Publishing Co. Ltd., Pp. 93-105.
- Baker, D.1989. Language Testing: A Critical Survey and Practical Guide, London: Edward Arnold.
- Carroll, B.J. 1980. Testing Communicative Competence: an Interim Study, Oxford: Pergamon Press Ltd.
- Davies, A. (ed). 1968. Language Testing Symposium: A Psycholinguistic Approach, London: Oxford University press.
- _____. 1982. "Criteria for Evaluation of Testing of English as a Foreign Language", in Heaton, J.B (ed). Pp. 11-16.
- _____. 1984. " Procedures in Test Validation", in Bozok, S. and Hughes,A. (eds), Proceedings of the Seminar Testing English: Beyond The High School. Istanbul: Bogazici University Publications, Pp. 83-91.
- _____. 1985. "Communicative Language Testing". Journal of Applied Linguistics (Annual Publication of the Greek Applied Linguistics Association), No 1: 22-23.
- _____. 1990. Principles of Language Testing. Oxford: Basil Blackwell Ltd.
- Dejenie Leta. 1990. "Achievement, Washback, and Proficiency in School Leaving Examination: A Case of Innovation in an Ethiopian Setting". Unpublished Ph.D.Thesis. University of Lancaster, United Kingdom.
- Finocchiaro, M. and S. Soko. 1983. Foreign Language Testing: A Practical Approach. New York: Regents Publishing company, Inc.
- Fischer, R.A. 1982. "Measuring Linguistic Competence in a Foreign Language". Language Teaching, Vol 15:207-217.
- Forrest, R. 1975. "Objective Examinations and the Teaching of English". English Language Teaching, Vol XXIX, No. 3:240-246.

- Gay, L.R. 1980. Educational Evaluation and Measurement Competencies for Analysis and Application. Columbus: Charles E. Merrill Publishing Company.
- Graham, J.G. 1987. "English Language Proficiency and the Predication of Academic Success". Tesol Quarterly, Vol. 21 No. 3:505-521.
- Heaton, J.B. (ed). 1982. Language Testing. Hayes, Middlesex: New Avenue Press.
- Henning, G. 1987. A Guide to Language Testing: Development, Evaluation, Research. Cambridge, Massachusetts: Newbury House Publishers.
- Henning, G. et al. 1981. "Comprehensive Assessment of Language Proficiency and Achievement Among Learners of EFL". TESOL Quarterly, Vol. 15, No. 4:457-466.
- Hughes, A. (ed), 1988 (a). Testing English for University Study (ELT Document 127), Hong Kong: Modern English Publications with the British Council.
- _____. 1988. (b), "Introducing a Needs Based-Test of English for Study in an English Medium University in Turkey", in Hughes, A. (ed). Pp. 134-145.
- _____. 1989. Testing for Language Teachers. Cambridge: Cambridge University Press.
- Johnson, F.C. and C. Kin-Ling. 1981. "The Interdependence of Teaching, Testing, and Instructional Materials", in Read, J.A.S. Directions in Language Testing. Singapore University Press. Pp. 277-303.
- Ministry of Education. 1983. "Major Considerations for Promoting English Language Teaching in Ethiopia Secondary Schools Based on the 1981/82 English Workshop Participants' Report", Addis Ababa: English Panel, Curriculum Department.
- _____. 1986. Evaluative Research of the General Education System in Ethiopia: A Quarterly Study, Addis Ababa.

- Gay, L.R. 1980. Educational Evaluation and Measurement Competencies for Analysis and Application. Columbus: Charles E. Merrill Publishing Company.
- Graham, J.G. 1987. "English Language Proficiency and the Predication of Academic Success". Tesol Quarterly, Vol. 21 No. 3:505-521.
- Heaton, J.B. (ed). 1982. Language Testing. Hayes, Middlesex: New Avenue Press.
- Henning, G. 1987. A Guide to Language Testing: Development, Evaluation, Research. Cambridge, Massachusettes: Newburry House Publishers.
- Henning, G. et al. 1981. "Comprehensive Assessment of Language Proficiency and Achievement Among Learners of EFL". TESOL Quarterly, Vol. 15, No. 4:457-466.
- Hughes, A. (ed), 1988 (a). Testing English for University Study (ELT Document 127), Hong Kong: Modern English Publications with the British Council.
- _____. 1988. (b), "Introducing a Needs Based-Test of English for Study in an English Medium University in Turkey", in Hughes, A. (ed). Pp. 134-145.
- _____. 1989. Testing for Language Teachers. Cambridge: Cambridge University Press.
- Johnson, F.C. and C. Kin-Ling. 1981. "The Interdependence of Teaching, Testing, and Instructional Materials", in Read, J.A.S. Directions in Language Testing. Singapore University Press. Pp. 277-303.
- Ministry of Education. 1983. "Major Considerations for Promoting English Language Teaching in Ethiopia Secondary Schools Based on the 1981/82 English Workshop Participants' Report", Addis Ababa: English Panel, Curriculum Department.
- _____. 1986. Evaluative Research of the General Education System in Ethiopia: A Quartly Study, Addis Ababa.

- Pearson, I. 1988, "Tests as Levers for Change (or Putting First Things First)". ESP in the Classroom: Practice and Evaluation - ELT Document 128, Pp. 98-107.
- Spolsky, B. 1985. "What Does it mean to Know How to Use a Language? An Essay on the Theoretical Bases of Language Testing". Language Testing, Vol. 2, No 2: 180-91.
- Tassew Zewdie et al. 1990. "Assessment of the Degree of Correlation Between the ESLCE Results and Four Years of Secondary School and University Academic Performance". Addis Ababa: Institute of Educational Research, AAU.
- Underhill, N 1982. "The Great Reliability/Validity Trade-off", in Heaton, J.B. (ed). Pp. 17-23.
- Weir, C.J. 1988. Communicative Language Testing. Exeter: A. Wheaton & Co. Ltd.
- Wood, R.1987. Measurement and Assessment in Education and Psychology: Collected Papers 1967 - 1987. London: The Falmer Press.