

ANALYSIS OF THE ENGLISH LANGUAGE TEST COMPETENCIES AND ITS USEFULNESS: A CASE OF NATIONAL EXAMINATION COUNCIL OF TANZANIA (NECTA)

Masatu Julius Malima¹ Bosco Thomson Mtani² and Ngowi, Scholastika Albert³

Abstract

Provided that English language testing at micro and macro level has been subjected to test construct- underrepresentation and construct-irrelevant, the study aimed at analyzing language test competencies and its usefulness: A case of National Examination Council of Tanzania (NECTA). Both Integrative Communicative Competence Model by Celce-Murcia et. al. (1995) and Bachman and Palmer (1996) model of test usefulness were used as the theoretical lens. To achieve the objectives above, a case study design of NECTA-Certificate Secondary Education Examination (CSEE) for English language were systematically selected whereby 2009, 2011 and 2013 national English Examinations were chosen for analysis. The researchers used two themes for the study namely: tester variability and construct- underrepresentation. Documentary review and content analysis were used as a tool for data collection and analysis respectively. The researchers found that tester variability is both condition and a matter of degree, provided that other language competencies such as linguistic competencies and strategic competencies were easy to be tested and better achieved compared to other areas such as discourse competencies and socio-cultural competencies which found to be difficult in achieving using paper-pencil tests unlike performative test. In case of test usefulness and test-under construction, the findings also showed that the test items remain uncertain and matter of degree regarding the coverage-ability, inclusions, authenticity and intergrativeness as well as predicative-ability. Other language competencies were easily assessed using paper-pencil or written production while those required oral production, interaction and meditation were easily ignored, for instance: speaking elements, listening elements and pragmatic competencies. However, what seem to be achieved in terms of communicative approach remains superficial drills of communicative approaches. It is recommended that since language testing is different from testing social science subjects, it is likely that both performative and pen-pencil test should be designed by NECTA to achieve the all competencies of the language as claimed in the national secondary English language curriculum.

Key words: NECTA, CSEE, test construct- underrepresentation, tester variability, and language competence.

1.0 INTRODUCTION

The question on the usefulness of examination validity in Tanzania has got a long history since the Arusha Declaration (1967) when Mwl. J.K. Nyerere questioned on the existing nature of Cambridge examinations used in Tanzania education system. However, the quality assurance in terms of curriculum and examinations had been executed by Tanzania Institute of Education (TIE) and NECTA since 1972 and 1993 respectively. NECTA has given an authority for setting, administering, marking, publishing and certification of both CSEE and Advanced Certificate Secondary Education Examination (ACSEE) following the Education and Training Policy of 1995 which gave its full mandate (TIE, 2007).

However, both CSEE and ACSEE have been in to place following the 1960s education system (7-4-2-3). At that time Tanzania mainland introduced the Higher Secondary Education and also

¹ Assistant Lecturer at Moshi Co-operative University (MoCU). Email: masatu152@gmail.com

² Assistant Lecturer at Moshi Co-operative University (MoCU). Email: boscomtani@gmail.com

³ Assistant Lecturer at Moshi Co-operative University (MoCU). Email: scholangowi@gmail.com

introduced Dar es salaam University College in 1961 and the reorganization of Primary and Middle schools in 1960. The education system meant seven years of primary schooling, four years of ordinary schooling, two years of advanced schooling and three years of first year degree courses. The examination system namely the final examination assessments respectively were as follows: Primary School Leaving Examinations, Certificate of Secondary Education Examination (CSEE) and Advanced Certificate of Secondary Education Examination (ACSE) and degree examinations.

Since English language testing today has been run under its defined programmatic and integrative approach, TIE which has been responsible for curriculum design and development since 1972, acts as a co-work with NECTA to ensure that both language teaching, learning and language testing respectively fit into the current theory of English language teaching and assessments. For instance, testing theories at both O-level (Form I-IV) and A-level (V-VI) should reflect the objectives and language competences required by the students in learning the language. Such objectives and competences emphasize on the communicative competences of the learner taking into consideration the integrative approach of both leaning and assessment of the language. Both O-level syllabus (TIE, 2010) and A-level Curriculum (TIE, 2007. Pg. 14) imply that students' competence should be defined in terms of grammatical competences, sociolinguistic competences, communicative skills competences, and discourse competences to enable them to communication locally, globally, in professional, academic and vocational fields. While scholars confirm the quality of the existing English language curriculum, for instance O-level syllabus which is found to achieve quality, continuity, autonomy, discussions and its relevant to situation (Shayo, 2011), the current study is interested to investigate the tester variability in language competences and the assessments of construct –underrepresentation of the test instruments.

1.2 Statement of the Problem

Test construct- underrepresentation and construct-irrelevant have been reported to be among the problems for test measurement practice (Shahamy, 1998; Uyasal, 2010; Namaghi, 2011; Juan and Weiping, 2002). Such scholars argue that : i) some tests meet administrative agenda and/or carry bureaucratic agenda; ii) teaching is done for testing not achieving target proficiency; iii) test lacks utility in terms of conceptuality - not fit to all culture (Rhetorical norms, logic reasoning, and argumentation styles) and predicative. They argue that these tests have both negative wash back or/ and positive wash back at both micro level (classroom level) and macro level (national level) of test measurement (Shahamy, 1998; Uyasal, 2010; Namaghi, 2011; Juan and Weiping, 2002). However, the available studies while agree in common to have been test construct- underrepresentation and construct-irrelevant, in Tanzania, local study attempted to study on the quality of the curriculum in terms of objectives and aims in-line with the teaching practices unlike the aspect of test usefulness at macro-level which has remained unstudied. In that case test analysis (test usefulness) is found to be a new area of study as far as NECTA experiences a very recent- paradigm of the new Communicative Competencies of English Language Curriculum (2005). Therefore, the study aimed to analyze the Language Test Competencies and its Usefulness using a case of NECTA.

1.3 Research Objectives

1.3.1 General objective

The general aim of the study was to analyze the language test competencies and its usefulness: A case of National Examination Council of Tanzania (NECTA).

1.3.2 Specific objectives

- (i) To analyze the tester variability of the language competencies of the test instruments used by NECTA.

- (ii) To describe the test- usefulness and construct-underrepresentation regarding language competencies of the test instruments used by NECTA.

1.4 Research Questions

- (i) What is the tester variability of the language competencies of the test instruments used by NECTA?
- (ii) Does the test consists of either test- usefulness or construct-underrepresentation regarding language competencies of the test instruments used by NECTA?

1.5 Significant of the Study

Provided that English language testing at micro and macro level has been subjected to test construct- underrepresentation and construct-irrelevant, the findings from the current study address the challenge by establishing tester variability and test usefulness practices so as to identify areas which need improvements. The findings also are in-line with the 2025 Sustainable Development Goals (SDGs) which advocate for educational quality in Tanzania and at global level. At the discipline level, the study contributes in the area of language assessments and evaluation by focusing on the area of tester variability and test usefulness.

2.0 LITERATURE REVIEW

2.1 Operational Key Term

Language test competencies

Language competencies has been defined differently by scholars, but what they share in common is that language competencies should be seen in terms of language domains namely, linguistic competencies, strategic competencies, discourse competencies, social and cultural competencies and pragmatic competencies (Hymes, 19972; Conale and Swain, 1980; Conale, 1983; Benchman 1990; Celce-Murcia *et. al.*,1995). They emphasize that when the learner is taught and tested in these domains, he/she is likely to have covered what is referred to as language competencies. However, the current studies agree with the definition and thus adopted and modified version to be used in the current study.

2.2 Theoretical Framework

The current study employed both Bachman and Palmer (1996) model of test usefulness and Integrative Communicative Competence model by Celce-Murcia *et. al.* (1995) used as the Lens for investigating the tester availabilities and test construct of communicative competences.

While Bachman and Palmer (1996) recognize the six elements of test usefulness such as reliability, construct validity, authenticity, interractiveness, impact and practicality as explained in the graphical representation by Chan, (1999, p.7), they believe that test validity is a matter of degree that is to what extent the test measures the target language competences such as reading, writing and speaking abilities of the students. The validity is also defined to be the extent to which the interpretation of the test scores is appropriate to the student's language ability. Therefore, the current study defines validity in terms of the extent to which the test item measures the target language ability and thus an analysis of the construct-underrepresentation. Bachman and Palmer add that such interpretations standards should be identified. While construct validity has been a unitary term or concept, other validities like content validity, concurrent validity, predicative validity are also identified.

The second model namely Integrative Communicative Competence model by Celce-Murcia *et. al* (1995) attempt to explain the test constructs communicative competences. This acts as the benchmark to test developers, in that case, NECTA, to achieve the current pragmatic and integrative approach to test constructions. The model attempts to save as a lens in investigating the tester variability in English language construction as far as the current study is concerned.

The model explains five elements of language communication competences as the modified theory from communicative competence by Hymes (19972), Conale and Swain (1980), Conale,(1983), Benchman (1990) to Celce-Murcia *et. al.*(1995). Celce-Murcia *et. al* (1995) divided competence into linguistics competencies, sociolinguistics, strategic competences, discourse competences and actional competences. Discourse competences concerns with logical sequencing of the text (spoken or written). The competence shapes the other competencies and vice versa.

Linguistic competences makes the basic elements of communication, they include sentence patterns, morphological system and lexical resources. Sociolinguistic competences refer to the speakers' knowledge of expressing appropriately in social and cultural contexts in which they are produced. An actionable competence is about the understanding of the speech acts from the speakers' communicative utterances. The model explains that the above competencies are influenced by strategic competences concerns with the knowledge of communicative competencies such as non-verbal cues and how to use them. While the model saves as guide for performative tests, the current study focuses on the analysis of the pen-pencil tests gets a limited operation by the model. However, the indirect performative test items will be taken into consideration to the model competencies.

2.3 Empirical Literature Review

The available studies present a mixed results on the quality and usefulness of language test instruments; while other test instruments have found the construct validity of the tests, others have found construct- underrepresentation and construct-irrelevant. They argue that these have both negative wash back or/ and positive wash back at both micro and macro levels (Shahamy, 1998; Uyasal, 2010; Namaghi, 2011; Juan and Weiping, 2002). While different tests may be found to test for specific language competences they have found themselves construct underrepresentation bearing in mind that such tests are designed to meet the administrative agenda from which the test carries its bureaucratic agenda for all its described purpose (Shahamy, 1998). She continues, teachers themselves found teaching for test not for target language proficiency as a case of English for Foreign Language (EFL) oral proficiency designed for secondary school leavers. Her study in 1993 further found that though the test consists of the role play, interview and Monologue, teachers taught only tasks found on the oral tests.

Uyasal's (2010) observations from the international standardized test namely IELTS brings a consideration to the current study as far as construct underrepresentation is concerns. He concerned questioned on the international writing tests for IELTS validity. He raised the questions that what should be called international standards in terms of rhetorical conventions, genres? Would the argumentation styles, logical reasoning, organizational patterns, rhetorical norms common across culture? He, thus, suggests further considerations on the quality and designation of IELTS. Further, Namaghi (2011) did a comparative study between test tasks and target use tasks with the aim to explore the possible construct underrepresentation and construct irrelevant difficulties in Shiraz University Language Proficiency Test (SULPT). He found that SULPT suffers from construct-underrepresentation and construct irrelevant difficulties. For instance, the reading tasks found difficulties reported by 80.4% of the students compared to its real life reading tasks. This meant that the test does not capture the important aspects of reading proficiency and therefore lacks predicative utility. The test had also negative wash back to students as they emphasized much on aspects which were well presented and downplayed those aspects that seem pose difficulties. He suggests for test developer to eliminate construct underrepresentation and construct irrelevant variances.

After a number of reforms in College English Test in China since 2005 following the heated discussion in terms of its test contest, the newly formed national English test for Chinese

college students claims both to measure communicative competences and better reflections of English proficiency of the college students, thus, a good predictive in their academic performance.

Chen (2009) did a comparative study between the old and the new College English Test Band 4 (CET-4) in china. While the study focused on the six qualities of the CET-4 tests using the test usefulness model by Bechman and Palmer (1996) he found the following: (i) The newly introduced test found with higher validity than the old one because the new CET-4 test explicitly identifies the constructs to be tested as well as how they can be tested. These findings concurred with Juan and Weiping (2002) who found the CET-test I with low validity; (ii) The new CET-4 tests covered a wide range of listening and reading skills by adopting a varied techniques of testing, iii) the grammatical structure and the vocabularies were tested in the context unlike isolation, thus achieved the current pragmatics and integrative testing approach; (iii) The score interpretation were valid as they reflected the learners' language ability in terms of their strength and weakness; (iv) In terms of authenticity and interactivity or remove authentic in a limited degree, however, the new test found more authentic than the old one in listening (reflected the daily life conversations), reading (had scanning and skimming that used in both academic and daily life setting) and writing part; and (v) In terms of wash back, the new CET-4 test motivated the students to engage in quantitative and qualitative learning because of the wide range of listening and reading tasks.

Therefore, while the above studies did not tell the tester availabilities in terms of test language competences measured, it is also the interest of the research to analyze the construct under representations in NECTA which seems to be a new area of study as far as the above studies are concerned. Following different observations above, the current study needs to analyze the construct underrepresentation and the tester variability in language competences for ASCEE and CSEE in Tanzania as far as the above study.

3.0 METHODOLOGY

3.1 Study Design

The study employed a cross sectional research design from which the data were collected once. To achieve such design the researcher employed analytical approach to establish the extent to which the test instruments were integrative into English language theories of testing and whether the test achieved the construct validity. Mixed approach of analysis were used such that the data were firstly reorganized into structure form by content analysis then turned into a fine form for quantitative analysis to get percentages of the test item occurrences.

3.2 Population, Sample Size and Sampling Techniques

The current study did not target the tester takers as the target population under study, rather it targeted the collections of English national examinations for ordinary form four done between 2008 to 2014. Purposive sampling was used to make decision as to why from 2008 to 2014. The reason was that the selected years were important as they lie within the communicative approach since 2008 of its commencements in English language teaching, learning and testing. Further, systematic random sampling was used to select the three tests for analysis namely 2009, 2011, 2013 National English Language Test for Form Four. The test structure, scoring distributions and sample item techniques are shown in table 1, 2, 3, and 4 in the appendix 1, page 16.

3.3 Definitions and Measurement of Themes

Tester variability was defined in terms of the following themes: linguistic competencies, discourse competencies, pragmatic competencies, strategic competencies and social and cultural competencies: So as to achieve such language competencies: the following measurement

indicators/sub-themes was used for analysis: Firstly, linguistic competencies- the variability of the test developer was measured in terms of the following sub-themes: test techniques and skill tested; secondly, it was the discourse competencies. The variability of the test developer was measured in terms of the following sub-themes: techniques used, the goal centre /the skill to be tested and the knowledge center; thirdly, it was strategic competencies-the variability of the test developer was measured in terms of the following sub-themes defined under reading skills: testing techniques used, the comprehension level to be tested and the strategies expected to be used by the test taker in high-stakes tests: Also the variability of the test developer was measured in terms of the following sub-themes defined under writing skills: writing genres, goal centre, the content knowledge required, the monitor required and executive process required; fourthly, social and cultural competencies was measured in terms of the following sub-themes: test techniques and skill tested/ability measured.

Test usefulness and construct-underrepresentation was measured using the following measurement indicators: degree of redundancy of the test items, degree of inclusiveness of the test takers, degree of complexity required/appropriateness to students' language ability, neglected competencies/coverage ability, predicative ability, wash-back, authenticity and intergrativeness, and grammar and structure tested in context and not in isolation.

3.4 Data Collection Tools

The research objectives were achieved using documentary review as instrument for data collections. Kothari (1986) explains that the method involves the review of any written material in that case English National Examination as shown in appendix part 1.

3.5 Data Analysis Techniques

The current study employed Content Analysis (C.A) suggested by GAO (1989) from the fact that the method turns unorganized materials into systematic materials for both qualitative and quantitative analysis. Moreover, since the nature of the study objectives needs a content analysis, the method through a step and procedural analysis addressed the issues of reliability and validity of the findings. The following were the steps used in qualitative data analysis: (i) At step one after the researcher decision to use the C.A, the English test tools (see appendix 1) were taken for analysis. While the test sections were taken as the context of analysis, the test item was taken as the unit of analysis to obtain the nature of test item in use as a recoding unit. After selecting the unit of analysis; (ii) the second step involved the coding of the categories by the use of matrixes. In that case, four categories were established such as linguistic competence of test items, social and cultural competence of test items, discourse competence and strategic competence of the test items a shown in appendix 1. Gao (1989) emphasizes that coding categories should reflect the objectives and the content in that case objective one and two of the current study. Such categories were also inclusive, exclusive and independent as suggested by Berelson (1952); (iii) Step three, the researcher treated each test item into their related coding categories by considering the nature of English test theories; their characteristics in relation to test items; (iv) Step four involved the coding of the materials manually; there were no electronically coding because what believed to be for instance, behavioral test items were implicative from the theoretical point of view of English language testing; and (v) Final step, the data were summarized by using frequencies as a scale measurement of which the findings were quantitatively analyzed and represented in a bar graph and tables (see, findings and discussion section below). An interpretation and establishment of the data validity were done using the findings and the available documents. To achieve the reliability of the test item analysis, all the coding were checked against independent coders and against the testing theory criteria under practices. The validity was checked against the existing related studies.

4.0 FINDINGS AND DISCUSSIONS

4.1 Introduction

The section attempts to answer the two specific objective above which focused on the tester variability and test-usefulness. The two variables were examined and discussed basing on the following: strategic competences, Linguistic competencies, Social-cultural competencies and Discourse competencies as the aspects of communication approaches. The nature of the test used for analysis and discussion were CSEE 2009, 2011 and 2013 as shown in terms of examination structure, score distributions and sample techniques used in table 1, table 2 and table 3 in the appendix part I.

4.2 Strategic Competencies

4.2.1 Tester variability

Strategic competencies influence the all remaining language competencies as it is concerned with the knowledge of communicative competencies such as non-verbal cues and how to use them. Strategic competencies were analyzed within the domains of both reading and writing skills: as for instance, the results in figure 1 shows that strategic competencies in reading skills found to be tested in both section “A” namely “comprehension and summary” in question 1 and 2 throughout the three examinations under study. In case of reading skills, strategic competencies were measured using the comprehension level to be tested and the strategies expected to be used by the test taker in high-stakes tests. The analysis showed that students were exclusively 100% tested for achieving local-interpretative competencies¹ in 2011 unlike 2009 and 2013. Moreover, global-interpretative² found to be tested at 50% in 2009 which is more than half of the other years. Global-literary³ was found to be in balance for all of the three years under investigation as the findings showed 33.33% across the three years. Moreover, other reading competencies found not to be tested at all for instance interpretative⁴ local and literary-interpretative⁵(see figure 1).

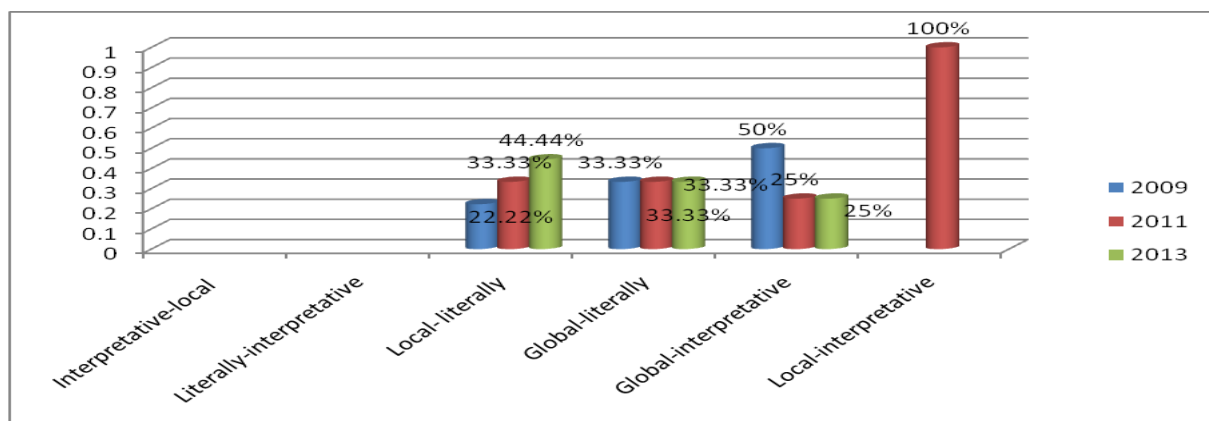


Figure 4. 1: Strategic Competencies-Reading skills tested in CSEE 2009, 2011, 2013

The study findings show that the strategic competencies in writing skills found to be tested in section “C” namely “language use” in question number 10a, b, c and d. The variability of the test developers were measured in terms of writing genres, goal centre, the content knowledge required, the monitor required and executive process required. The findings are shown in table 4 in the appendix part.

The findings showed that the tester used the following writing genres: academic essays, friendly letter, and speech writing across the three years (2009, 2011, 2013). The purpose of the test items were to measure the students’ ability in imaginative writing, problem solving skills,

writing for opinions, suggestive skills, argumentative skills, evaluative skills, narrative skills, speech writing and invitation writing skills. However, the students required to use the following knowledge: imaginative, input processing, schema/socio cultural background, creativity and imaginative with maximum or minimum use of monitor requirement. The learner required to use executive process in terms of planning, organizing, generating, translating, evaluating and revising and editing skills (see table 4 in the appendix part).

4.1.2 Test usefulness and test-under reconstruction

In case of strategic competences the tester included the important aspects that could predict students' ability in writing responses from which most of writing test items demanded executive process in terms of generating, organizing, planning, revising, editing, translating and evaluating. In case of level of complexity, the reading materials were all at the level of the learners provided that there were at the level of their syllabus. To some extent the tester achieved the balance in case of test variability regarding of global-literary competencies, as it was found to be in balance for all of the three years under investigation as the results showed 33.33% across the three years. This implies that, the test tested the student's ability to comprehend within the level of the sentences, paragraphs and words as well as enabled them to use their pragmatic knowledge and schema to respond to test items. However, in case of reading test items, the summary and comprehension test items used by the tester found not to achieve inclusiveness; for instance, students were found not to be tested in other reading competencies like interpretative local and literary-interpretative. Other competencies were exclusively 100% tested while leaving other reading competencies. Moreover global-interpretative found to be tested at 50% in 2009 which is more than half of the other years. The analysis showed that students were exclusively 100% tested for achieving local-interpretative competencies in 2011 unlike 2009 and 2013.

4.2 Linguistic competencies

4.2.1 Tester variability

Linguistic competences makes the basic elements of communication, they include sentence patterns, morphological system and lexical resources. Linguistic competences variability of the test developers were measured in terms of test techniques used and skill tested. The findings from content analysis in table 5 below shows that linguistic competencies is tested in section "B" in question number 3, 4, 5, 6, and 7 under the section namely "Patterns and vocabulary". Such observations are consistent throughout the three examinations under investigations. The findings showed that tester used a variety of testing techniques to achieve linguistic competencies such as a list of related words with word not a member, a list of short sentences with incomplete inflectional morphology, a list of optional words with a filling black short sentences, sentence with wrong spelling, forms of word in pair to be filled in blank spaces with the right form and other test techniques as shown in table 5. The aim of the test items were to measure a variety of skills namely; associations, word formation (inflectional form of the word), the basic meaning of the word, the written form of the word (orthography), grammatical behaviour of the word and register of the word. Other abilities measured were an ability to use subordination conjunctions, the use of hypothetical sentences, accurate use of prepositions, ability to report indirect speech, use of subordination conjunctions, ability to construct complex sentence, ability to use personal adjective, ability to use cases and relative pronouns and ability to use question tag, time, condition sentence and subordination.

Table 5: Linguistic Competencies

Examination year	No	Testing techniques	Skills tested
2009	3	A list of related words with one word not a member	association
	4	A list of short sentences with incomplete inflectional morphology	Word formation (inflectional form of the word)
	5	Sentence with long spelling	The written form of the word (orthography)
	6	A list of optional words with a filling black short sentences	The basic meaning of the word
	7	Forms of words in pair to be filled in blank spaces with the right form	Grammatical behaviour of the word
	3	Incomplete sentences with a given instructed alternative responses	The grammatical behaviour of the word
	4	Sentences with incomplete spelling	The written form of the word (orthography)
2011	5	Matching list words to its appropriate basic vocabulary	Association
	6	Given a short family text with short questions related to family vocabulary	The register of the word
	7	List of sentences with instructed response	Ability to use subordination conjunctions, the use of hypothetical sentences, accurate use of prepositions
	3	Transformational task and instructed response tasks	Ability to report indirect speech, use of subordination conjunctions, ability to construct complex sentence, ability to use personal adjective
	4	Lists of optional words to be filled in a blank short sentences	Ability to use cases and relative pronouns
	5	Word in a sentence with wrong incorrect spelling	Written form of the word (orthography)
	6	A list of occupational related vocabularies to be filled in its appropriate blank expressions	The basic meaning of the word (denotation)
2013	7	Sentence with instructional response	Ability to use question tag, time, condition sentence and subordination

4.2.2 Test usefulness and test-under reconstruction

In case of coverage-ability, the tester provided enough forms of grammar and structure that where not beyond their level of complexity provided that there were within their English language syllabus. The tester achieved wash back provided that the test engaged students in both qualitative and quantitative learning because of the wide range of grammar and structure

testing materials. Moreover, grammars were not tested in isolation provided that stimulus were included within the sentences, list of words and paragraphs.

However, what seems missing is that in communicative approach structure and grammar testing are needed to be within the context of communication but not to be tested alone as separate grammatical drills. What seems to be done by the tester is what Murton (1960, 1966) asserted, that the mind of the learner should be kept out of meaning till later where he /she will use the patterns as the tool for communication with more structural patterns being learned at the earlier stage. The main assumptions of behaviorist in language teaching and testing (Bloom field, 1933; Zlling Harris, 1957; Chomsky, 1957; Lado and Fries 1957) were to keep the individual learner into mastery of the inventory of phoneme, morphemes and patterns of syntactical operations; meaning was treated autonomous to grammar.

4.3 Socio-Cultural Competencies

4.3.1 Tester variability

Sociolinguistic competences refer to the speakers' knowledge of expressing appropriately in social and cultural contexts in which they are produced. Speech acts are also reported to be part of language use in real context. For instance, Cutting (2002) identifies these speech act as an utterances which are intended to promise, apology and threaten (speech act theory), the unit also introduces critical discourse analysis an approach which analyses the purpose of the language in the social context.

The findings in table 6 show that socio-cultural competencies were found to be tested in section "C" namely, "language use" in question number 9 consistently in the three examinations under investigation. The test developer availabilities reported in terms of testing techniques and the ability measured. The tester used limited testing techniques such as matching item through- out the three examinations under investigation. The ability measured was also limited to grammatical behaviour of the sentence in its appropriate contexts, appropriate use of the words in its context of use.

Table 6: Socio - cultural competencies

Examination year	No	Testing techniques	Ability measured
2009	9	Matching item	The grammatical behavior of the sentence in its appropriate context
2011	9	Matching item	Appropriate use of the word in its context of use
2013	9	Matching item	Appropriate use of the word in its real context

4.3.2 Test usefulness and test-under reconstruction

Socio – cultural competencies were found to be less reflecting the real life conversations, as the tester used a limited test items in that case, matching items and even the target skills measured did not reflect the real life conversations as they look like grammatical drills. However, this might also have contributed by the nature of the exam which is pen-pencil paper unlike performantive test. Scholars assert such questions are superficial communicative drill test items which are worse, and meaningless to student in an attempt to use the language in real life conversations (Paulston & bruder, 1975).

This section confirms to the findings that language live in the social context where the role of interactions (cultural relevance) takes place (Brown, & Yule. 1983). The invalidity of the test items in these sections are the irrelevant items which measures the student 'WHAT' consists of

language use and language skills i.e recall, comprehensions, instead of 'HOW' to perform and use the language in its real environments, as in case of matching items in question 9 which seems superficial communicative drills.

The questions lack authenticity and integrativeness, as for instance, how to test for communicative competences and situational competences were treated unfairly. The test items were organized the same to strategies used in testing other subjects like History, Geography and General Studies. English unlike other subjects is the discipline of its own with its own theories of learning, methods of teaching, and testing. For instance the tester should be able to differentiate the testing of words as grammatical drill and word usages in its real context of use as used in question 9 throughout the three examinations under investigation.

The tester also found to be unfair by neglecting some of the language competencies such as speaking, listening, speech acts, and pragmatic competencies which are found to be important part of language competencies, thus lack of predicative-ability of the test in terms of students' evaluation and certification regarding English language competencies.

4.4 Discourse Competencies

4.4.1 Tester variability

The findings as shown in table 7 provide that discourse competencies found to be tested in section "C" namely "language use" in question 8 for all of the tests selected under investigations. The findings showed the test developer variability in terms of techniques used, the goal centre or the skill to be tested and the knowledge centre. The findings from content analysis showed that discourse competencies measured in terms of language use were evidenced by using lists of sentences with illogical sequences with the goal to achieve students' logical flow of the paragraph.

Table 7: Discourse Competencies

CSEE Years	Test techniques	Goal Centre/skill tested	Content knowledge
2011, 2009, 2013	Lists of sentences with illogical sequences	To achieve the logical flow of the paragraph	Contextual knowledge

4.4.2 Test usefulness and test-under reconstruction

Some of the test items found to be biased provided that when tester required to test for students' schema, the schema found to favour the scientific major students as in case of CSEE (2011), this is in case of discourse competencies, thus lack of inclusiveness. The test items seem not to have predictive-ability as skill tested is limited to coherence within the paragraph unlike the cohesion beyond the paragraph and coherence devices. Throughout the exam the test items found to be of redundancy provided that the same test techniques were used throughout the three examinations under investigations which may turn the teaching to be for testing and not for mastering English language proficiency (Shahamy, 1998).

5.0 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

The researcher found that , tester variability is a matter of degree and conditional, provided that other language competencies such as linguistic competencies and strategic competencies were easy to be tested and much achieved compared to other areas such as discourse competencies and social and cultural competencies which found to be difficult in achieving using paper-pencil tests unlike performative test. In case of test usefulness and test-under construction, the test items remains uncertain and matter of degree regarding the coverage-ability, inclusions, authenticity and integrativeness and predicative-ability. However the problem in assessing

English language competencies found to be bias-based since the use of paper-pencil/written production found to achieve limited English language competencies while ignoring other competencies which required oral production, interaction and meditation. Moreover, what seem to be achieved in terms of communicative approaches by NECTA remain superficial drills of communicative approaches.

5.2 Recommendations

Since language testing is different from testing social science subjects, it is likely that both performative and pen-pencil test should be designed by NECTA to achieve the all competencies of the language as presented by scholars, in that cases, strategic competencies, linguistic competencies, pragmatic competencies, discourse competencies and actional competencies. Moreover, NECTA should make sure that test variability achieves the testing usefulness principles adopted (coverage-ability, inclusions, authenticity and intergrativeness and predicative-ability) in this paper and from other scholars.

REFERENCES

- Alduais, S.M.S.A. (20102). An account of the approach to language testing. *International journal of Academic research in progressive education and development*, 1(4), 42-58.
- Browm, G., & Yule, G. (1983). *Discourse analysis*. New ork: Cambridge University pres
- Tanzania Institute of Education, (2010). English Language Syllabus for Secondary Schools Form I-IV. Dar es salaam, Tanzania.
- Cueto, S. *et.al.* (2009). Pschometric Characteristics of cognitive development and achievement instruments in round two of young lives. *An international studies of childhood poverty*.
- Cutting, J. (2002). *Pragmatics and discourse: A resources book for students*: London and New York: Routledge.
- Kluitman, S. (2008). *Testing English as the foreign language two EFL- testing use*. Germany: Albert-ludwigs- university Freiburg
- TIE (2007). Curriculum Development for Advanced Secondary Education in Tanzania: Ministry of Education and Vocational Training. Dar Es Salaam-Tanzania: Institute of Education (TIE).
- Martinez-Flor, A., & Uso-Juan, E. (2008). Teaching intercultural communicative competence through the four skills. *Revista alicantine de. estudios ingleses*, 21(2008), 157-170
- Philippson, L.H.R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.
- Selinker, .L., & Susan. M. (2008). *Second language acquisition: An introductory course*. London Routledge.
- Shayo, E. (2011). New Syllabus-2005-English Language for Secondary schools, Form I-IV. EzineArticles.com, <http://EzineArticle.com/expert/Euge-Shayo/896688>.
- Wood, B. B. (2001). Stake's countenance model: Evaluating an education professional development course, *The journal of environment education*, 32(2), 18-27.

APPENDIX 1**Table 1: SCEE 2013 English Language Examination structure, Score Distributions and Sample Techniques**

Section	No	Nature of the task	Technique used
Section A- Comprehension and Summary (10%)	1	Comprehension text	-Multiple choices
	2	Short text	Summarizing task
Section B (20%) Patterns and Vocabularies	3	Pattern structure	Transformational task and instructed response tasks
	4	Pattern structures	List of optional words to be filled in a blank short sentences
	5	Vocabulary	Words in a sentence with wrong incorrect spelling
	6	Vocabulary	A list of occupational related vocabularies to be filled in its appropriate blank expressions
Section C (30%) Language use	7	Pattern structure	Sentences with instructional responses
	8		List of sentences with illogical sequences
	9		Matching item
	10		a) Narrative essay, b) Application business letter, c) narrative essay on wedding d) invitation card and an outline essay
Section D (40%) Response to readings	11	Poem text	Both comprehension questions and evaluative questions
	12	Poem and prose	Differetiative (essay) essay
	13	Plays	Suggestive response essay from four optional plays
	14	Novels	Validation response essay from six optional novels

Table 2: SCEE 2011 English Language Examination structure, Score Distributions and Sample Techniques

Section	No	Nature of the task	Technique used
Section A- Comprehension and Summary (10%)	1	Comprehension text	-Multiple choices and instructed responses
	2	Short text/passage	Summarizing task
Section B (20%) Patterns and Vocabularies	3	Pattern structure	Incomplete sentences with a given instructed alternative responses
	4	Vocabularies	Sentences with incomplete spelling
	5	Vocabulary	Matching list words to its appropriate stem vocabulary
	6	Vocabulary	Given a short family text with short questions related to family vocabularies
Section C (30%) Language use	7	Pattern structure	Lists of sentences with instructed response
	8		List of sentences with illogical sequences
	9		Matching item
	10		a) A composition essay b) speech writing, c) argumentative essay, d) speech

Section D (40%) Response to readings	11	Poetry related response questions	Discursive response essay from three poems
	12	Text of poem	Comprehension questions from poem text
	13	Plays related response questions	Supportive response essay from four optional plays
	14	Novels related response questions	Discussive response essay from six response essay

Table 3: SCEE 2009 English Language Examination structure, Score Distributions and Sample Techniques

Section	No	Nature of the task	Technique used
Section A- Comprehension and Summary (10%) Section B (20%) Patterns and Vocabularies	1	Comprehension text	-Multiple choices and fill blanks
	2	Short text/passage	Summarizing task
	3	Vocabulary	A list of related words with one word not a member
	4	Patterns	A list of short sentences with the incorrect verbs
	5	Vocabulary	Sentences with wrong spelling words
	6	Vocabulary	A list of optional words with a filling black short sentences
	7	Vocabulary	Form of words in pair to be filled in blank spaces with the right form.
Section C (30%) Language use	8		List of sentences with illogical sequences
	9		Matching item
	10		a) Expressive essay, b) opinion essay c) opinion essay
Section D (40%) Response to readings	11	Two stanza poem	A comprehension and evaluative short questions from the two stanza poem
	12	Poetry related response questions	Discursive response essays from three optional poetry
	13	Novel related response questions	Supportive response essay from four optional plays
	14	Plays related response questions	Supportive essay

Table 4: Strategic competencies-writing skills tested in CSEE 2009, 2011 and 2013

Year of exam	NO	Writing genres	The goal centre; purpose	The content knowledge required	Monitor require	Executive process required
2009	10a	Essay	Testing on imaginative essay writing	Imaginative	Minimal monitor	-planning, organizing, generating translating revising, evaluating for imaginative essay

	10b	Friendly letter	Testing on friendly letter writing	Input processing	Minimal monitor	Organizing, generating translating, evaluating and editing for friendly letter
	10c	Essay	Test on problem-solving essay writing	Schema/social-cultural background, creativity	Maximum monitor	Student required for Generating, organizing, translating, evaluation and editing for problem-solving essay writing
	10d	Essay	Testing on opinion/suggestive essay writing	Social-cultural background	Maximum monitor	Student required for generating, organizing, translating, evaluation and editing for suggestive essay writing
2011	10a	Essay	Test on evaluative essay writing	Creativity and input process	Maximum monitor	Students required for generating, organizing, translating, evaluation and editing for evaluative essay writing
	10b	Speech writing	testing on speech writing	Creativity and imaginative	Maximum monitor	Students required for generating, organizing, translating, evaluation and editing for the speech
	10c	Essay	Testing on argumentative essay writing	Imaginative and creativity	Maximum monitor	Students required for generating, organizing, translating, evaluation and editing for argumentative essay writing
	10d	Speech writing	Testing on speech writing	Imaginative, and creativity	Minimal monitor	Students required for Generating, organizing, translating, evaluation and editing for speech
2013	10a	Essay	Testing on	Creativity	Maximum	Students required

		narrative essay writing	and imaginative	m monitor	for generating, organizing, translating, evaluation and editing for narrative essay
10b	Friendly letter	Testing on friendly letter writing	Input processing	Minimum monitor	Students required for generating, organizing, translating, evaluation and editing for friendly letter
10c	Essay	Testing on narrative essay writing	Imaginative and creativity	Minimal monitor	Students required for generating, organizing, translating, evaluation and editing for narrative essay writing
10d	Invitation card and telephone message	Testing on invitation card and outlining the format for telephone message	D (i) Input processing, and creativity, D (ii) input processing	Minimal monitor	Students required for generating, organizing, translating, evaluation and editing for invitation card and outlining the organization of telephone message