

The Impact of Test Content Validity on Language Teaching and Learning

Dr. Ahmed Gumaa Siddiek

Dawadami Community College - P.O. Box 18

Shaqla University – KSA 11911

E-mail: aahmedgumaa@yahoo.com

Abstract

This paper investigates the impact of content validity of language tests on both teacher and learner. Educational evaluation produces too much stress in both teacher and learner, but it is given less attention by the teacher than any other teaching tasks. This inconvenience comes from the doubt with which this topic is approached. Tests are supposed to focus on core syllabus constituents to find how much the learner achieved. In this attempt we went through some examinations that seemed to lack content validity. We found that when the exam lacked this feature, most of the teaching would be focused in the techniques of examination rather than doing real teaching. It was also found that students did not refer to their textbooks as source of knowledge but exerted all efforts only to collect marks. The result of these practices yielded failure in the achievement of pedagogical objectives of language education.

Keywords: Language testing, Content validity, Test comprehensiveness, Backwash, Language education

1. Introduction

Educational assessment is the responsibility of teachers and administrators not as mere routine of giving marks, but making real evaluation of learner's achievements. Teachers are the frontiers who are assigned to carry out the policies of education by teaching specific syllabus, within a limited time span to achieve specific pedagogical communal or individual goals. Most discussions about the value of educational assessment focus on its use for the purposes of grading the students' marks and ignore other more important role of helping students with their learning. Cliff: (1981:17) believes that, assessment must be used to help students. It can operate in two ways: Firstly, any teacher knows that assessment, particularly in the form of examinations, will determine to a large extent when students will study, what they will study, and the effort they will put into their work. Secondly: when students are provided with a programme of progressive assessment with appropriate information as to their strength and weaknesses, they can take action to correct and improve performance prior to any major or definitive assessment.

Although there is too much literature about tests and testing, the issue is still a highly neglected area by many language teachers and by some test designers. This inattention is due to lack of knowledge of testing techniques and lack of unawareness of the importance of testing among the teachers community. One small glance on teacher-based examinations would betray too many facts about the poor performance of some teachers in designing and structuring classroom tests.

From an unstructured questionnaire administered by the writer among a group of college staff, he came to a surprising fact that the majority of the interviewees had very little idea about key terms used in the testing area such as: test validity, content validity, reliability, proficiency test, achievement test, taxonomy of educational objectives, high-stake examinations, item analysis, standardized test, comprehensive tests or backwash of tests. Ignorance of such terms was/is of great negative consequence on the teaching and the learning process as well. This unawareness was due to the fact that most of the staff (subjects) in the interview had not been exposed to courses in testing at first degree level. A little percentage acquired some knowledge while they were doing further studies in the master's programmes and a very small portion got some knowledge from their free reading. Due to this poor command of testing techniques, the writer could detect too many major faults with their exams. It was evident that teachers took very little time in preparing their tests. The subjects were not able to make simple analysis to provide reports of their students' achievements. They were also unable to interpret the meaning of figures of the test scores because they lacked the basic statistical knowledge. This kind of knowledge

is still not included in most of the BA, B.Sc programmes, although it is of great importance when used in the analyses of social phenomena. Some statistical techniques can help teachers, educators and decision makers to understand and interpret the results of examinations in scientific technical terms. Statistical knowledge should be a core part of any research or course in ELT fields.

2. Theoretical framework and previous literature

The researcher carried out a survey in Juba University - Sudan in testing intended to investigate the degree of attainment of the educational objectives as tailored by the Ministry of general Education in the teaching of the English language programs in the secondary schools in the Sudan. The subjects were M.ED students. Their job was to analyze and classify the test items according to a specific criteria based on the Bloom's Taxonomy of Educational Objectives. To my dismay the subjects were not familiar with Bloom's taxonomy of Educational Objectives; hence, they were *unfamiliar* with such terms as cognitive domain, affective domain, psychomotor domain- although they were BA holders in English and had some long teaching experience in secondary school education. So, I had to give them some training in the *Taxonomy of Educational Objectives* and explain the different level of educational *classes* of the cognitive domain, i.e. knowledge, comprehension, application, synthesis, analysis and evaluation. I had to provide them with a list of the taxonomy of critical thinking as applied to writing tasks of verbal actions of the six classes of knowledge. It was intended to enable them classify the items according to the action verbs of the testing of mental abilities.

Another study was also carried out by the writer in the evaluation of the Sudan School Certificate English Examinations. Those examinations qualify students to university admission. They were found as proficiency tests rather than achievement tests, supposed to test students' language progressive learning and not their overall language proficiency skills. See Siddiek: (2004)

The researcher did not come through any research that investigated the language testing in his country. Testing traditions is a neglected area in almost all Arab educational institutes. But it has just recently begun to draw some attention in Saudi Arabia, where the researcher gained long teaching experience to the young Saudis. It is also a neglected area in Sudanese institutes but new programmes are being introduced by Nile Valley University to MA in ELT two or three years ago. It had also been taught for sometime to M.ED students at Khartoum University in the College of Education.

Although the Department of Educational Evaluation, in the Ministry of Education is responsible for carrying out continuous evaluation of the (SSC) examinations in general and in English Examinations in particular, the researcher did not come through such work that is analyzing items of examinations of the English language final examinations,

2.1 Statement of the problem

In his research, Siddiek:(2004) found that, the Sudan school certificate English examination in that form was not a standardized achievement test as it did not embody the features of the achievement tests. It was also found that the Sudan School Certificate English examinations (SSCE) in that format were incomprehensive: that it did not cover most of the materials which was supposed to be taught to students' in their textbooks. This incomprehensiveness yielded the recent situation of lack of content validity of the school English Examinations, which we are going to investigate in this paper. Those tests did not reflect the real performance of the Sudanese students in English and also they lacked backwash - which is the effect of exams on the teaching and learning processes.

2.2 Objectives and Audience

This paper is intended to shed light on some of the characteristics of a good test and to find to what extent those features are present and applicable in the English exams in the Sudan. So the paper tries to investigate the pitfall and shortage within these exams as well as suggesting some remedies and recommend some point for reform in Language education. This paper targets the English teachers' community, test designers, language learners as well as curriculum designers and education decision makers.

2.3 Methodology

A critical and analytical approach was used by going through the examination samples, and then inferences will be made from the general discussions and interpretations of the work.

2.4 Tests and Testing

According to Sax: (1980:13) a test is a task or series of tasks used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes. As a teacher you might have known many

people who are not happy with the schools burden their children with too much testing. It is easy to find an answer to such questions, but surely there are some answers more suitable and eloquent than others. Learners at schools need to be tested because:

- We as teachers want to promote students' knowledge, skills and cognitive abilities.
- Tests are facilitating factors that can help teachers deal with the individual student.
- Results of tests can help teachers modify their teaching strategies if the old ones did not work.
- Test results can help students select the fields of future especially which will suit their aspiration and fit with their abilities.
- Tests results can also help us give the right orientations to young learners.
- Test results are benchmarks and tools of quality control with which decision makers can secure the national educational goals.
- With testing reliable measurement we can detect the achievement of the totality of the pedagogical aims on personal individual and communal base.

2.5 Language Testing and Language Learning

Williams:(1986:142) criticizes the view of teachers towards language testing when he says that teachers see this job as an area of English language teaching that many teachers shy away from. It is frequently seen as a necessary evil. But educational evaluation is an important process through which we can determine the success of our educational programmes and secure our educational goals. [It] includes any means of checking what the students can do with the language. Lado: (1975:20)

Assessment is an integral part of any effective teaching programme, thus it should be subject to planning, designing, modifying and frequent revision of its validity as tools of quality measurement. We assess our programmes to see how much success we have made and consequently take decisions according to the validity and reliability of those measurements. But assessment is often given less attention by the teacher than any other task, although it is an activity that arouses anxiety and stress in both teacher and student. The major source of this anxiety and inconvenience comes from the uncertainty with which the task is dealt with. We will set up from this point to handle the topic by asking what is assessment and how can it be done?

Schools examinations are sometimes felt to be a source of threat and discomfort for both teachers and students. They are often seen as tedious and boring, but they should be non-threatening and non-boring because a test which is felt that way by teachers and pupils, often leads to lower achievement. So, what information do we need in planning the assessment programmes and make it unthreatening and fruitful? We will need to answer the following leading questions as Hughes: (1995:1) put them:

- What is the ultimate goal of the test?
- Do we want to measure skills?
- Do we want measure knowledge?
- Do we want to know the subjects attitudes towards one specific phenomenon?
- Do we want to realize the extent of skill of proficiency in certain area of knowledge?

Then we need to answer more questions such as:

- What is the best time to carry out this task?
- What reliable and valid procedures do need to use to elicit the right behaviour from the learner?
- How to report the results of the elicited behaviour.

Every test is a learning situation. The results of tests must produce positive backwash, which is the effect of testing on teaching and learning. Backwash can be harmful or beneficial. [When] the test content and testing techniques are at variance with the objectives of the course, then there is likely to be harmful backwash. Hughes: (1995:1)

2.6 What to assess?

According to Cliff: (1981:27), aimlessness is the most single cause of ineffectiveness in teaching and of frustration of education effort. This is particularly true in regard to assessment; how can we as teachers construct a test or examination, or set an assignment, unless we have been able to learn which skills we wish the student to acquire. Therefore, if we want our assessment to be valid, it must measure what we have set as objectives of the

course. These objectives may be only transferring of a body of knowledge, skills, such as problem solving or to develop certain values in the students or aiming to change certain behaviour in the learner. Our task then in preparing a test is to write questions which cover all the material taught and which measure the course objectives. To enable teachers to match their questions to their teaching objectives, several educators have provided taxonomies of educational objectives such as Bloom (1956) in his (Cognitive Domain) in (1964) and (The Affective Domain): Scriven (1967) in categories of learning, also Gagne and Briggs (1974). These taxonomies have proven to be of great use to educators, test developers, syllabus writers as well as decision makers. The Bloom Taxonomy of Educational objectives were and still of great use and of great effect in education development in the USA and somewhere else in the world. The taxonomies put teachers and learners on the right track of achieving clear educational objectives, within the right time and at the right place, since both learners and educators share the same basic concepts and basic knowledge.

2.7 Identification of course content

Identifying the content of the course is the first step in the process of assessment. The test questions should cover all topics of the course with appropriate weighing of those topics. Educational Objectives as stated by Benjamin S. Bloom and others could be adopted as a basic yardstick in assessment. The theory of educational taxonomy dealt with the cognitive objective through a set of six major classes:

- | | | |
|--------------|------------------|----------------|
| 1. Knowledge | 2. Comprehension | 3. Application |
| 4. Analysis | 5. Synthesis | 6. Evaluation |

These classes can serve both the teacher and the learner by giving both the framework and the plan with which they can design their pedagogical goals and at the same time be able to measure the degree of the attainment of these educational programmes in measuring the development of mental abilities. The first two classes of knowledge and comprehension is said to test the lower ability of the learner as they are concerned in testing his/her memory of recalling facts and specifics. Whilst, the other four classes of the mental abilities of the taxonomy are expected to test the ability of the learner in critical thinking that needs high skills in using the brain effectively through inferences, deductions and interpolations

The taxonomy has also extended to cover the affective objectives, having to do with the attitudes, values, interests, and appreciation. The major five classes in the taxonomy of the affective domain are:

- | | | | | |
|---------------|---------------|------------|-----------------|---------------------|
| 1. Receiving. | 2. Responding | 3. Valuing | 4. Organization | 5. Characterization |
|---------------|---------------|------------|-----------------|---------------------|

A rapid change is taking place in the field of assessment and evaluation. Different techniques are implemented to help educationists, policy makers and decision takers to carry out their policies on scientific and statistical bases and that would help them achieve their objectives in the best way. We need to follow these procedures in our schools in the Sudan as we are aspiring to achieve *certain* goals in the future. These goals should be set in forms of syllabus and courses and be taught to the young generation to secure them the minimum education. Therefore, if we wish to achieve these goals we need to *precisely* define our educational objectives and *efficiently* carry them out. Assessment can help us attain our school programmes and help us disclose the weaknesses in the performance of our students, our teaching and the ineffectiveness of our syllabus, and then we will be able to develop learners' *overall* abilities and skills of language acquisition.

2.8 Characteristics of a good test

A good test must have the following features:

- It must be reliable.
- It must be valid.
- It must be comprehensive.
- It must be practical & scoreable
- It must have positive backwash

2.8.1 Reliability

Generally, there are three techniques for carrying out the assessment properly: oral test, written test and the practical tests as it is found in applied sciences where tests are run in laboratories. But in all cases the test must be reliable. Understanding the term *reliability* according to Sax (1980:255) is very important for two reasons: first, principles of test constructions depend upon having a clear understanding of this term. A test of low reliability is a waste of time for both teachers and students since it permits no conclusion to be drawn. Second: the selection of test depends, in part, on a consideration of the reliability of measurements. Unreliable tests are

no better than assigning students random scores. Thus a good test is a reliable test and reliability is defined by Sax (1980:257) as the extent to which measurement can be depended on to provide consistent, unambiguous information. Measurements are reliable if they reflect "true" rather than chance aspects of the trait or the ability measured. A good test- therefore-is a reliable test; a test that measures consistently. On a reliable test you can be confident that someone will get more or less the same score, whether they happen to take it on one particular day or on the next; whereas on an unreliable test the score is quite likely to be different, depending on the day on which it is taken. Hence, reliability of a test is the extent to which it is free from random measurement error. Tests that are highly reliable are usually consistent and trustworthy, Hughes: (1995:22)

2.8.2 Validity

Technically, validity can be calculated as Sax: (1980:258) put it as true variance divided by obtained variance. In practice, true variance, of course, have to be estimated since it can not be computed directly. We can have some more definitions of validity. "A test is valid when it measures what it claims to measure." Garrett. (1964:30). "A test is valid when it "measures what it ought to measure." Ebel: (1972:436.) "Does the test measure what it is intended to measure? If it does, it is a valid test." Lado: (1975:30) "A test cannot be a good test unless it is valid. The essence of validity means the accuracy with which a set of test scores measures what it claims to measure." Abbott: (1992:178.) "A test is said to be valid if it measures accurately what it is intended to measure" Hughes: (1995:22)

A valid test is a tool on which we can absolutely depend to give consistent measure and sound judgment. A ruler is a valid tool but it is *impractical* to use in measuring the distance from Khartoum to Nyala, although it can give the same reading all the time it is used; so the ruler is a reliable tool but it is not a valid tool in this specific situation.

2.8.3 Content Validity

According to Hughes: (1995:27), a test is said to have content validity if its contents constitute and represent sample of the language skills, structures, etc. with which it is meant to be concerned. A test is said to have face validity, if looks as if it, measures what it is supposed to measure. A valid test is the test which affords satisfactory evidence of the degree to which the students are actually reaching the desired objectives of teaching, these objectives being specifically stated in terms of tangible behavior. Content validity is an important notion for measurement of skills and knowledge that are frequently employed in evaluation studies. In this context, content validity usually refers to the *degree* to which a measure captures the program objective (or objectives.). Anderson: (1975:460) refers content validity to the extent to which the test we're using actually measures the characteristics or dimension we intend to measure. Chair: (2003) thinks that content validity refers to the extent to which the test questions represent the skills in the specified subject area. Content validity is often evaluated by examining the plan and procedures used in test construction. He adds that content validity is the accumulation of evidence to support a specific interpretation of the test results.

Validity is traditionally subdivided into three categories: content, criterion-related, and construct validity, Brown: (1996, 231-249). *Content validity* includes any validity strategies that focus on the content of the test. To demonstrate content validity, testers investigate the degree to which a test is a representative sample of the content of whatever objectives or specifications the test is originally designed to measure

When the outcome of a standardized test is used as the sole determining factor for making a major decision, it is known as high-stakes testing. Common examples of high-stakes testing in the United States include standardized tests administered to measure school progress under No Child Left Behind (NCLB), high school exit exams, and the use of test scores to determine whether or not a school will retain accreditation. These tests are supported by some, especially politicians, who believe that schools need more accountability. However, the practice of high-stakes testing is heavily criticized by many parents and educators, who believe that the outcome of a standardized test should be only one of many things taken into account when reaching a major decision about education.

2.8.4 Comprehensiveness

It is another quality of a good test. A good test must be comprehensive to cover all or almost all the aspects of the course contents which are embedded in one specific syllabus. The comprehensive test must as well attain the educational objectives of the target course or the unit taught to students.

2.8.5 Practicality and scoreability

Another quality of a good test is to be practical in its layout and management in term of time consumption, human physical and mental efforts and money-wise economical as well as being easy to score and report its

results. The reporting of the test is a very important process as the learner is waiting to see his result which he will make use in his practical life.

2.8.6 Washback or backwash

The term is used to mean the impact which the test is going to make on both the teacher and the learner. The washback can be positive to yield good influence on both the teacher and learner. It can also of negative effect and yield bad results on the teaching and learning processes. Absence of fairness and lack of content validity or incomprehensive plus the invalid ways of rating of tests all together are factors that generate negative influence on the share-holders whether they are teachers or learners, parents, decision makers or employers.

3. Specifications of Educational Objectives

Assessment cannot be achieved unless we have clearly stated the educational objectives that the student must reach. Goals and objectives must be written appropriately to ensure that they are valuable and feasible. It is important to interpret these objectives into concrete behaviour, which can be seen and observed. But this process may face some problems, Tyler, in Ross:(1963:114) summarizes them by saying all methods of measuring human behavior involves four technical problems: (1) defining the behavior to be evaluated, (2) selecting the test situation, or determine the situation in which the behavior is expressed, (3) developing a record of the behavior, and (4) evaluating the behavior recorded. But how can we avoid such problems and construct reliable and valid tests and avoid the inaccuracy in our tests? As we see there are two components of test reliability: the first component is the performance of the students in their different examinations and the second component is the way teachers mark or rate the examinees' performance. Hughes: (1995:36) suggests some ways of achieving consistent (reliable) performances from the candidates by:

- taking enough samples of behaviour: the more items that you have on a test, the more reliable that test will be. It has been demonstrated empirically that the addition of further items will make a test more reliable. And each item should as far as possible represents a fresh start for the candidate.
- not allowing candidates too much freedom: In some kinds of language test there is a tendency to offer candidates a choice of questions to allow them a great deal of freedom in the way that they answer the questions that they have chosen. This much freedom would expand the range over which possible answers might vary.

The researcher has run such an experience while marking students' performance in one (SSCE) Examinations. The candidates used to range too far in their answers by writing too many lines beating round the correct answer; while one word would have been sufficient to give the right response. This behaviour was due to the number of staves (space) given to the candidate to record his response.

3.1 Standardized Tests and High-Stake Examinations Evaluation

Standardized tests can be used to measure participant skills. They are so named because their administration, format, content, language, and scoring procedures are the same for all participants these features have been 'standardized'. Standardized tests according to Anderson: (1975:384), attempt to provide for measurement of individual differences in as unambiguous ways as possible. Thus, the process of standardization permeates all aspects of testing: construction, administration, scoring, reporting, and interpretation of results.

These examinations are sometimes locally developed such as the Sudan School Certificate Examination (SSCE); or commercially created in most achievement areas and for some aspects of language proficiency such as the international American test of English known all over the world as TOEFL, and the British English IELTS. When considering the definition of "standardized test," it is clear that all high-stakes tests (should be standardized). The term standardized test originally meant, and still means when used precisely, a test that:

- Has been carefully, expertly constructed with analysis and revision;
- Has explicit instructions for uniform (standard) administration; and has table of norms (standard) for score interpretation derived from administration of the test to a defined sample of students.
- "Loosely, the term can refer to almost any published test or inventory, whether standardized in the manner just describe or not." Ebel:(1972:465)

3.2 Three categories of standardized tests: Proficiency tests

These tests are designed to measure the over-all peoples' ability in a language regardless of any prior training they may have had in that language. The content of a proficiency test therefore, is not based on the content or the objectives of language courses which people taking the test may have followed. It can be taken by any one who

wanted to take that test or who happened to be there at the moment of the test. So Proficiency tests do not need prerequisite requirements.

3.3 Achievement Test

These tests are directly related to language courses. Their purposes are to establish how successful individual students, group of students, or the courses themselves are. Scholastic tests which students take at the end of school year represent the greatest part of these tests. These tests are designed to help in placement of students in school classes or colleges and universities such as the (SSC) examinations. They are standardized test developed to measure skills and knowledge learned in a given grade level, usually through planned instruction, such as training or classroom instruction. Achievement tests are often contrasted with tests that measure aptitude, a more general and stable cognitive trait. Achievement test scores are often used in an educational system to determine what level of instruction for which a student is prepared. High achievement scores usually indicate a mastery of grade-level material, and the readiness for advanced instruction. Low achievement scores can indicate the need for remediation or repeating a course grade. (http://en.wikipedia.org/wiki/Achievement_test)

3.4 Diagnostic Test

As the name suggests, it is the test that is used to identify the weaknesses and strengths of a students in certain language elements such sounds, intonation, stresses, morphemes and skills such as writing and reading.

There are other types of tests such as the placement test, promotion tests that. And there is the remedy test which is used to help select those students who need special help and treatment so as to do remedial work to treat their weakness at certain area of knowledge. Subsequently we need not to say, these functions of tests apply to all subjects including the foreign language. According to Al-Khuli: (1995:95), the same test may perform several functions simultaneously. Thus a test that measures student's achievement may be used by the teacher in self-evaluation as well. It may also be used as a promotion criterion to provide patterns with information about their children.

4. Variables and Strategies of Language Testing

4.1 The elements of language

Lado: (1975:25), divided these variables into two categories: Language is built of sounds, intonations, stresses morphemes, words and arrangements of words having meaning that are linguistic and cultural. The degree of mastery of these elements does not advance evenly but goes faster in some and slower in others. Each of these elements of language constitutes a variable that we would like to test.

4.2 The Skills

We have four more variables to be tested, namely the degree of achievement in speaking, understanding, reading and writing. A fifth skill is the ability to translate, which should be tested as an end in itself and not as a way to test mastery of the language.

4.3 Stages of Test Construction

Test construction is a practical problem for many teachers. It is a real challenge which they frequently face. It is a skill that teachers should train themselves in, in the best accurate manner so as not to do their students any harm, of course - most of the time unintentionally. But what is the best way to do that? According to Hughes: (1995:49) we have to ask the following questions when constructing a new test what kind of test is it to be? Achievement, final or progress, proficiency, diagnostic or placement test? What are its precise purposes? What abilities are to be tested? How detailed must the results be? How accurate must the results be? How important is the backwash?

4.4 Preparing a table of Specifications

Analysis and planning of the content of an achievement test is very crucial. To represent the achievement test for any educational material, we must analyze the content of this educational material to see how much of truth the specific test shows by calculating and then determining the relative weight on the basis of time spent in the teaching of each subject. The identification of the relative weights of the educational objectives of the subject depends on the experience of the teacher. It is often the active and knowable expert teacher we need to consult in determining these weights depending on the objectives to be measured. Achievement tests should be representative sample of educational outcomes and the content of educational material. Practically, we can not put all the questions that we want to ask in the test at the end of the semester, where there are hundreds of facts and terminology we expect learners to know within the time allowed; but we can insert an appropriate number of paragraphs in the test to measure parts of the facts or applications

4.5 The purpose of the table Specifications

The most important purpose is to achieve a balance in the test, and make sure that the measurement of the samples is highly representative of the objectives of our teaching, and the subject content that we want to measure in the achievement test, to secure the ultimate goal of education. Educational goals embody the mental, physical and spiritual growth of the human individual. It also constitutes of the aim to change the individual person attitudes positively by engineering behaviour to secure pedagogical objectives.

4.6 Benefits of the Table of Specifications

- It helps build a balanced test with the size of efforts in teaching each topic.
- It gives real weight of each part of the subject, and therefore every subject deserves questions according to its relative importance.
- It helps in the in the selection of a representative sample of the objectives of teaching, in an orderly manner. The process measures the extent of an achievement to a large degree, the teacher makes the distribution of questions at different levels to those goals.
- It gives the student a lot of confidence of test fairness, which will assist the candidate in the organization of his time.

4.7 Component of the table of specifications

The table must consist of the following components:

- The relative weight of the subject and vocabulary that will be measured by student achievement performance.
- The relative weight of the objectives to be measured. The relative weights will determine the number of questions contained in the test of each topic and each level of the cognitive objectives.
- The relative weights to determine the question or questions for each level of objectives to be covered by the test.

4.8 Principles be taken into account when building a table of Specifications

We must consider the following principals when writing a table of specifications:

- the nature of the subject, and the educational set of behavioral objectives
- The length of time it took to teach each topic of the course.
- The nature of students with regard to the level of study.

4.9 Backwash and Lack of Content Validity

Backwash is an important concept in the field of applied linguistics. It refers to the impact and influence that the test can have on the teaching and the learning process. This influence can be positive or negative. The backwash concept is very much connected with the test content validity. If the test lacks the content validity it consequently yields negative backwash on both the teacher and the learner.

5. What happens if the test lacks content validity?

So when writing achievement test items, writers must begin with a list of content standards (either written by content specialists or based on state-created content standards) which specify exactly what students are expected to learn in a given school year. The goal of item writers is to create test items that measure the most important skills and knowledge attained in a given grade-level. The number and type of test items written is determined by the grade-level content standards. Content validity is determined by the representativeness of the items included on the final test.

(http://en.wikipedia.org/wiki/Achievement_test)

Content validity –representing of sample of the language skills, structures, etc- is a very important factor in the success of test. When all the language elements and skills are included in the test it will put pressure on the learner to cover the whole syllabus to read all the materials in his syllabus. It also makes teachers focus to teach the specific material within the specific course in the specific time. Content of test puts the learner and the teacher on the right track by committing themselves to the syllabus which is especially designed by experts to secure the educational objectives of the individual and the community.

5.1 The Sudan school Certificate English Examinations & Content Validity

The Sudan School Certificate English Examination (SCC) lacks content validity according to Siddiek (www.2010) who found that most teachers were concerned with the teaching of the testing techniques rather than of teaching the syllabus. Students did not refer to their textbooks as they thought it was waste of time because the examination would not be from those textbooks. Many teachers were discontent and unhappy with the structure of the (SSC) English examinations in its recent form. They said that most questions were irrelevant, most of the time extracted from old sources such as (THE NILE COURCE).

This irrelevancy made students *refrain* from reading and referring to their textbooks. Students strongly believed that no reading or summary passages would be selected from that specific (SPINE 4, 5 or 6) textbooks. There were also too many complaints of the incomprehensiveness of questions; that questions did not cover most of the syllabus which students had studied in their final secondary school year. Thus the test constituents were almost not representative of the samples in the syllabus. Therefore the behavioural objectives which were interned to be elicited by testing those young people were very much limited or almost nil. Examiners, who were assigned for constructing such high-stakes examinations, should have examined how closely a test reflects the curriculum and the standards of the examinees.

Those examinations were expected to show the degree of students' achievements in the course that they had read and not their language *overall* proficiency skills, a point that can be supported by Menke:(1998) when he pointed out that, in order for test to be effective and useful for educators in instructional practice, they must be deeply entwined with the classroom teaching and learning driven by the standards. It is also supported by Bronwyn: (2002) who said, if tests are aligned with standards and curricula, students will have an increased chance of demonstrating what they know and are able to do.

Albussairi :(2009) found that students did not read (Spine) Textbooks and they did not benefit from their English programmes, this is because the test did not frame the learners' behaviour towards his syllabus. So when the test lacks content validity, it means that the items are not covering the constituents of the syllabus and consequently the educational objectives of the syllabus will not be attained and the whole process of teaching will be for gaining nothingness. Most of the teachers' efforts were directed to help students pass the exams and collect the marks and no real teaching took place. Textbooks were neglected and not read so this would make money spent for nothing. You can just imagine a book which is written and not meant to be read, so why were thousands of copies published and distributed to students by the Ministry of Education? Why did we have to pay money for publishers, experts and teachers, hence the textbooks would not be used and referred to by the targeted audience who were in this case the secondary school students in the Sudan.

5.2 A model Exam

A test is said to have face validity, if looks as if it, measures what it is supposed to measure. Tyler in Ross (1963:114) regards a valid test as the one which affords satisfactory evidence of the degree to which the students are actually reaching the desired objectives of teaching, these objectives being specifically stated as in terms of the kind of behavior. A quick glance at any on of the test can easily betray the way the test was designed see (<http://moe.gov.sd/exams/2008/academy/english>). Scholastic tests are supposed to be achievement tests i.e. to test the specific amount of knowledge and skills imbedded in the textbook rather than being a proficiency test that tests the overall ability of language skills and knowledge of the learner without sticking to one specific programme or unit.

The (SSCE) as found by Siddiek (2010) was a proficiency test rather than an achievement test; therefore it did not test the material in the textbook but focused on a wider-angle of general knowledge in English skills. The test in its form is similar to TOFEL or IELTS, because it did not focus mainly on the issues of the syllabus, but it dealt with the *overall* language abilities and skills.

For example if we take 2008 exam, see (<http://moe.gov.sd/exams/2008/academy/english>) as a sample we would see details trying to test different language structures. The whole models could have been taken directly from the passages and pieces in the textbook of (SPINE 6). The summary and the vocabulary test could have been directly extracted from the textbook which was rich with such material. The question on the grammar section was testing only one grammatical aspect: suffixes and prefixes, totaling about 40% of the general language structures. It is a pity to realize that this question was based on only one page on the textbook although it represented 20% of the total marks of the examination.

If the test focuses on the textbook, the teacher will be obliged to teach the details in that textbook and the learner will be obliged to read and study the material in that textbook. When both the teacher and the learner focus their

effort on teaching and studying the specific programme, we will be then sure about securing our educational goals, because we will be sure of attaining the minimal education which we have embedded in our syllabus. The curriculum in the school or college is the minimum education that we think our learners should acquire within a limited time span and at specific mental age. This is what we are really targeting to help develop the overall mental and affective growth of young learners.

A test with no content validity would spoil the whole educational process. It would spoil the teaching as teachers would focus *all* their efforts to teach the *techniques* of answering these questions on the exam paper rather than doing *real teaching* to provide the learners with TRUE language situations which could be of practical use in their individual communications and in their future jobs.

These examinations also would spoil the learners as *all* their learning strategies would be directed to *rehearse* the right answer to the exam items. They would make no real studying of their textbooks because those books were practically of no use since the examination would not cover the material in them. Learners would be *exam-oriented hunters* rather than *learners* seeking language knowledge and language skills.

Thus these examinations in their recent forms lack the content validity because the constituents of the test are unrepresentative samples of the whole parts in the textbook. They did not cover the syllabus by testing its constituents and thus did not achieve the aims in the textbooks which were intended to provide the learners at this level of education with the minimum structured and planned programmes of English language to secure the ultimate of teaching English as a foreign language to the young Sudanese learners at secondary school.

6. Conclusion

It is not that difficult to teach a unit or course to your student, but it is important to know how successful that teaching is. The success of that course or unit can be seen through the right assessment of the efforts exerted to teach the course. Good teaching consequently yields good testing. The good testing is supposed to yield a positive backwash on both the learner and the teacher simultaneously. In examining the Sudan School certificate English Examinations (SSCEE) we came to find that those examinations lacked the characteristics of a good English Language test. Through his work Siddiek (2004) found that those examinations were proficiency test rather than standardized scholastic achievement scholastic tests. They lacked comprehensiveness, content validity and did not possess any positive backwash on the learners and teachers. This incomprehensiveness and lack of content validity yielded very poor educational practices in term of foreign language teaching in our country. This situation can be seen in this context:

- Teachers focus on teaching the examination techniques to help students collect as much marks as possible rather than doing real teaching to change the students' behaviour towards language learning.
- Students did not refer to their textbooks but focused their efforts in learning how to do the exam rather than learning the language to use in real life situations.
- The educational aims were not secured by the teaching of the syllabus imbedded in the textbooks. The result of this practices yielded poor attainment of national and personal pedagogical and educational goals.

7. Recommendations and Suggestions

Further studies are recommended in the field of item analysis of the (SSCEE). Immediate reform strategies of examination should take place, since examinations -among other things - are tools of quality control through which we can judge the success of our pedagogical and educational objectives. Examinations can put both the teacher and the learner on the right track of the whole education process. They can tell the teacher the degree of success or failure of his teaching methods, so he can change to new ones if the old ones did not work well. Examinations can also help the learner adapt and modify his learning strategies to gain real knowledge which can be of practical use in his life career. Examinations as well are benchmarks and yardsticks that can help decision makers to see how much of their national, regional or global education goals are achieved and secured.

References

- Al-Khuli, Ali Mohammad. (1999). Teaching English to Arab Students.
- Anderson, Scarvia. B. (1975). *In Encyclopedia of Educational Objectives*. Jossey Bass Publishers. San Francisco. California. USA.
- Bloom, Benjamin S et.al. (1956). *Taxonomy of Educational Objectives. Book 1. Cognitive Domain"*. Longman Group Limited, New York.
- Bronwyn, Carole. (2003). [Online] Available: www.cal.org.ericcl/Digest/subject/hhttp# assessment.

- Brown, J. D. (1996). [Online] Available: http://www.jalt.org/test/bro_8.htm.
- Cliff, John C. and Bradford W. Imrie. (1981). *Assessing Students, Appraising Teaching*. Grom Helm ltd. London SW 11David McKay Company. Inc. New York. 176-1851979.
- Ebel, Robert. (1972). *Essential Educational Measurements*. Prentice Hall, Inc. Englewood, Cliff. New York.
- Garrett, Henry. (1959). *Testing for Teachers*. American Book Company. New York. Glossary of Measurement Terms. (2003).
- Hughes, Arthur. (1995). *Testing for Language Teachers*. Seventh Printing. University of Cambridge. Bell & Bain, Ltd. Glasgow.
- Lado, Robert. (1975). *Language Testing*. Wing Tasi Cheung Printing Company.
- McNamara, Tim. (2001). *Language Testing*. Oxford University Press.
- Menke, K in Orlich, Donland.C, et.al. (1998). *Teaching Strategies*. 5th edition. Houghton Mifflin Company, Boston New York. U.S.A.
- Orlich, Donland. C, et.al. (1998). *Teaching Strategies*. 5th edition.
- Sax, Gilbert. (1980). *Principles of Educational and Psychological Measurement and Evaluation*. Second Edition, Wadsworth Publishing Company. Belmont, California. USA.
- Siddiek, Ahmed Gumaa. (2001). Analysis & Evaluation of Sudan School Certificate English Examination. A survey based on the Bloom's Taxonomy of educational objectives (a term paper as part of an M.ED), Juba University. Khartoum-Sudan.
- Siddiek, Ahmed Gumaa. (2009). Assessment of Sudan School Certificate English Examination. Published PHD thesis. [Online] Available: www.Lulu.org 2009. UK.
- SPINE SERIS (4,5,6). (1998). The National Centre for Curriculum and Educational Research. Ministry of Education. Sudan Secondary School Certificate Technical Manual 2003. Ministry of Education. Khartoum. [Online] Available: www.moed.gov.edu/sudan/exam/.
- Tyler. (1963.114). In Ross, C.C. Revised by Julian C. Stanley. *Measurement in Today's Schools*. 3rd edition. Englewood Cliff .J Prentice – Hall.