European Journal of English Language Teaching
ISSN: 2501-7136
ISSN-L: 2501-7136
Available on-line at: www.oapub.org/edu

10.5281/zenodo.59148 Volume 1 | Issue 2 | 2016

TEACHING TO THE TEST: ASSESSMENT SYLLABUS AND VALIDITY OF ENGLISH EDUCATION IN BOTSWANA SECONDARY SCHOOLS

Fetogang, Ernest B.
Educational Foundations, University of Botswana

Abstract:
The demand for better results among school leavers in Botswana, like in many other places in the world, has been increasing in recent years, but what do the scores students get in examinations really mean? Do they mean such school leavers are adequately equipped to join the world of work and fully participate in the development process? Is classroom instruction geared towards achieving the goals of the syllabus or it is a mere examination preparedness exercise? To what extent can mid-year and final examinations be taken as a valid measure of the syllabus?

To find the answers to these questions and others, the study, which was survey inferential in nature, comparatively content analysed the subject matter content as well as cognitive skill content of the midyear as well as the final examinations, and the subject matter content as well as well as cognitive skill content of English language syllabi. The data was then analysed using SPSS by conducting Pearson correlation and Fisher's z-score transformations tests. Both literature review and findings for this study indicated a tendency among instructors to teach more to fulfil the requirements of the examinations than those of the syllabus. Correlation between the content of the mid-year examinations and that of BGCSE was found to be significant in most cases. It was recommended that teachers be engaged in in-service courses that emphasize test development and that assessment of English should include oral exams.

Keywords: assessment syllabus, validity of education, teaching to the test, examinations

1. Introduction

In a bid to make the curriculum more relevant in Botswana, Botswana government introduced various programs as indicative of its long term vision of 2016. However, in recent times, the field of education in Botswana, has been criticized for producing human resources in large numbers that are lacking practical skills that could enable them to perform at tertiary schools and ultimately at work. This is due in part to the fact that classroom instruction is increasingly becoming examination oriented as opposed to being geared towards developing those skills that will enable them to fully participate in the long term development of the country.

Lee (1990) state that human resources development in Africa, provided mainly by government schooling, universities and vocational education, is aimed at formal employment and is overproducing skills for a stagnant labour market. Important skills needed to develop and guide programs are lacking at many levels, especially policy analytical skills.

It is unfortunate that there is little proof that these examinations are actually aligned to the syllabus, it is unfortunate that there is little proof that these examinations are actually aligned to the syallabus, and this makes the validity of inferences that these are made from their results questionable, as Kober (2002) has given that any form of teaching to the test is questionable if is raises students examinations scores without also increasing their knowledge and their skills in the broader subject being measured.

In Botswana, almost all syllabuses show a decline in performance at Grade C or better. Out of a total of 28 senior high schools, all show a decline in performance measured with number of candidates obtaining Grade C or better except about only 5 schools. For example, the total number of candidates with 5 C's or better decreased from 69,827 in 2010 to 65,303 in 2011, a decrease of 6.48% (see Table 1). The proportion of candidates reaching the standard required for Grade C or better in all core subjects decreased by 0.25% from 9.47% in 2010 to 9.22% in 2011. Apart from Home management, Music and Fashion and Fabrics, generally, there is decline in performance across all subjects. (Botswana Examinations Council, 2011).

Botswana has a history of high stakes assessment with an emphasis on the use of examination results to judge the quality of students and that of schools. Unfortunately, when test results are used in such minimal ways, the end result is that testing drives instruction as well as what counts as learning, hence narrowing the indicators of effective and efficient education system (Isaacs, 2007). Assessment often times has a negative impact on teaching and learning as it limits student-centered learning. This discourages educational innovation and does little to encourage the use of assessment

for refocusing teaching in order to improve the quality of education offered (International Reading Association (IRA), 2005). Similarly, an emphasis on defining school performance in terms of success on national tests and alignment with the national curriculum seems to attract a full range of local initiatives (Ntuane, 2005). Instead of reforming, assessment ends up driving instruction, thus reducing the very stronghold of an education system which is to equip learners with a fuller range of robust skills for future opportunities as they face the changing world and also to meet community development needs. This also means that as much as we want the examinations to be curriculum based, the curriculum also need to be assessed to be up to date with the fast growing needs of the society.

Bloom taxonomy provides a way to arrange topics and locate their level addressed during teaching and learning with the curriculum as a reference and that without referencing to the latter, it will not be easy to remain focused.

Baker (2001) states that in many educational and psychological measurement situations, there is an underlying variable of interest. This variable is often something that is intuitively understood, such as "intelligence." When people are described as being bright or average, the listener has some idea as to what the speaker is conveying about the object of the discussion. Similarly, one can talk about scholastic ability and its attributes, such as getting good grades, learning new material easily, relating various sources of information, and using study time effectively. In academic areas, one can use descriptive terms such as reading ability and arithmetic ability. Each of these is what psychometricians refer to as an unobservable, or latent, trait. Although such a variable is easily described, and knowledgeable persons can list its attributes, it cannot be measured directly as can height or weight, for example, since the variable is a concept rather than a physical dimension.

A primary goal of educational and psychological measurement is the determination of how much of such a latent trait a person possesses. Since most of the research has dealt with variables such as scholastic, reading, mathematical, and arithmetic abilities, the generic term "ability" is used within item response theory to refer to such latent traits (Baker, 2001).

The primary role of the test characteristic curve in item response theory is to provide a means of transforming ability scores to true scores. Under item response theory, the primary purpose for administering a test to an examinee is to locate that person on the ability scale. If such an ability measure can be obtained for each person taking the test, two goals can be achieved. First, the examinee can be evaluated in terms of how much underlying ability he or she possesses. Second, comparisons among

examinees can be made for purposes of assigning grades and awarding scholarships (Baker, 2001).

An & Yung (2014) narrate that item response theory (IRT) is concerned with accurate test scoring and development of test items. You design test items to measure various kinds of abilities (such as math ability), traits (such as extroversion), or behavioural characteristics (such as purchasing tendency). Responses to test items can be binary (such as correct or incorrect responses in ability tests) or ordinal (such as degree of agreement on Likert scales). Traditionally, IRT models have been used to analyze these types of data in psychological assessments and educational testing. With the use of IRT models, you can not only improve scoring accuracy but also economize test administration by adaptively using only the discriminative items.

In formal education the curriculum, as operationalized by the syllabus, takes a centre stage in the processes of teaching, leaning and assessment (see Figure 1). Given the intention of education, these processes are valid to the extent that they are syllabus-driven. Because of lack of adequate relevant skills teachers oftentimes resort to examination-driven teaching and assessment in their classrooms.

According to Fetogang (2015), in Figure 1, the *A* relationships representing the validity of public and school examinations, should be strong and highly significant while the *B* and *C* relationships should be weak and random. Teaching and assessment should all look at the curriculum for content and objectives and not at the content of past public examinations. Learners are handicapped when teaching and classroom assessment are examination-driven because they are only exposed to what could be easily measured through high-stake examination items, thereby narrowing the curriculum at the wish of assessment.

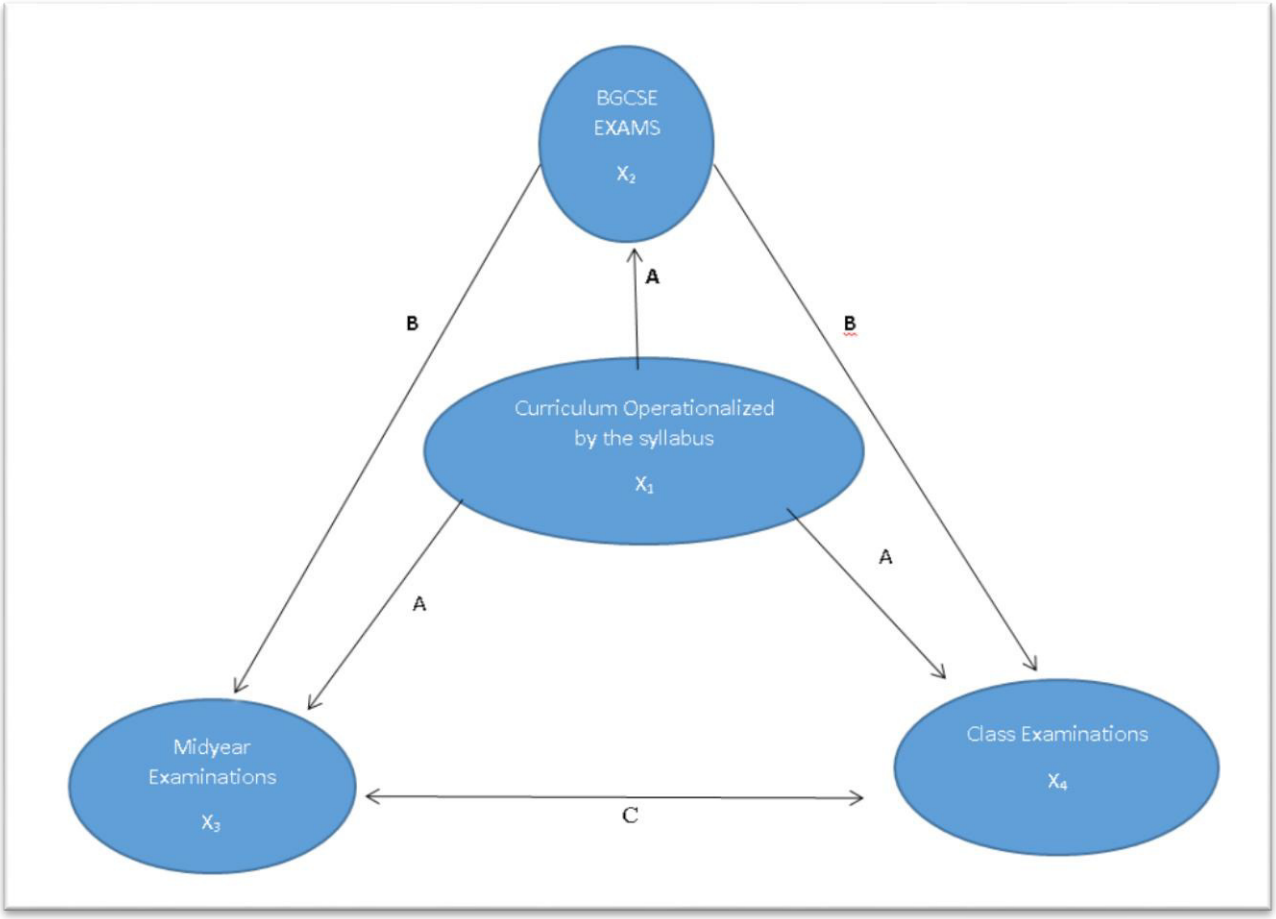

Figure 1: Syllabus-examination relationship in practice (Adopted from Fetogang, 2015)

Definitions and Abbreviations

Validity of Public Examinations - the degree to which the contents and skills measured by BGCSE exams reflect the content and the skills demanded by the Botswana curriculum.

Public exam - An examination offered by a national or provincial (state) authority, or on behalf of such an authority, to students at a particular level of an education system. The primary purpose is to certify the level of achievement of individual students and/or to select students for the next level of the education system.

Examination syllabus - A document formally specifying what will be assessed by the examination and how the assessment will be carried out.

BGCSE examination - mean Botswana General Certificate for Secondary Schools Examinations of which students undergo learning for 2 years.

Mid-year examinations – national examinations taken by all students in preparation for the final examinations

School-based assessment - Any assessment of student performance that takes place in a school and is incorporated into the public examination result. **CSBPS** - Cognitive skill and/or behaviour as provided for in the syllabus
CSBMMYE - Cognitive and/or behaviour as measured by mid-year examinations
CSBMBGCSE - Cognitive skill and/or behaviour as measured by BGCSE
SMCPS - Subject matter contents as provided for by the syllabus
SMCMMYE - Subject matter contents as measured by mid-year examinations
SMCMBGCSE - Subject matter contents as measured by BGCSE

Problem and purpose of the study

Juma (2005) states that it is true that pupils are drilled to pass examinations. He points out that drilling is worse in public boarding schools. The political leaders, tribal leaders, religious leaders, parents, the media and other government leaders expect student to perform well and that has not been the case in Botswana Senior Secondary schools for the past several years. Table 1 shows performance of students in English language in 2011 compared to previous years. There has been a significant decline in performance in the subject for the past years in Botswana senior secondary schools.

The study aims to review the extent to which the knowledge tested in public examination schools matches that which is expected of the student by the national curriculum. The study highlights the extent to which categories of both skill and content of school based English examinations as well as final examinations are syllabus driven.

Table 1: Performance of Candidates in English Language

Year	2007	2008	2009	2010	2011
Grade C or better	23.74	18.54	18.81	17.31	17.41

Research hypothesis

Ho_1: The validity of measuring English content by midyear and by BGCSE examinations with the content of the syllabus, as the criterion is not significant.

Ho_2: In the measurement of content, the validity of English by mid-year examinations given the English syllabus as the criterion is not significantly higher than the validity of

the English mid-year examinations given the English BGCSE examinations as the criterion.

Ho_3: Given the English syllabus as the criterion, the validity of midyear and BGCSE examinations in terms of the measurement of cognitive skill is not significant.

Ho_4: In the measurement of cognitive skill, the validity of English by mid-year examinations given the English syllabus as the criterion is not significantly higher than the validity of the English mid-year examinations given the English BGCSE as the criterion.

Literature review

Teaching to the test is a teaching is an approach taken with the express objective of modelling classroom instruction and assessment to the demands of the examination and nothing beyond it. The researchers found that on average, Teachers spend most of the time each week teaching reading followed by mathematics and writing, the non-tested subjects were taught on average much less frequently (Jones, Jones, Hardin, Chapman, Hargrove and Davis, 1999).

Jones, et al (1999) also found that when the majority of the instructional time is dedicated to a few tested topics, the rest of the time becomes a fought over community, subjects that are not tested are moved, removed and reduced as schools struggle with an overloaded school day. Sadly, these subjects commonly considered unimportant are the ones that complete the development of a child as a social being as they allow students to explore, interact and socialize. Mnthali (1995) assessed the teaching and examining of literature in English in Botswana and a questionnaire was used to gather data from students and teachers and an interview guide from education officers. His findings concluded that literature examinations had no content validity in the opinion of both teachers and students. A study by Jacob and Levitt (2003) aimed to identify cases of outright cheating on the part of the teacher or administrators on standard tests by engagement of teachers in such illicit activities as changing students' responses, providing students with answers or illegitimately obtaining copies of the examination before the examination is done. The researchers developed an algorithm for detecting teacher cheating that combines information on unexpected test scores fluctuations and suspicious patterns of student responses in a classroom.

It is worth noting that as in Akiyama (2004) it seems to be the case that the assessment of speaking would perhaps more importantly threaten deeply seated underlying values, though not expressively stated, which the present exam is assessing. In their own teaching, they paid little attention to speaking as it was not included in the

highly competitive university admissions tests that formed a primary goal of teaching at senior high school. Therefore, senior high school teachers did not recognize speaking skills as important in their context altogether (Akiyama, 2004)

From literature, teaching to the test seems like a universal problem. Elder (1997) investigated the use of scores from language examination. Most of the literature suggests that there are serious discrepancies between the syllabus content and that of national examinations. The study done by Sweing, Crisp, Ahmed and Pollity (2002) shows that normally, students develop certain expectations relating to examination questions, based on their experiences of classroom tests and past examination papers. A study by Cotton (1993) has indicated that employers are getting increasingly dissatisfied with the inadequacy of skills young job applicants have displayed in recent years. They may possess knowledge required but lack employability skills.

Nkosana (2008) presented and discussed the attitudes towards and perceptions of education officers regarding introducing a school-based continuous assessment of speaking in the Botswana General Certificate of Secondary Education (BGCSE) ESL examinations as provided for in the BGCSE English syllabus in Botswana. Three categories of education officers that are responsible for developing and evaluating syllabuses, teaching supervision (quality assurance), and assessment of the BGCSE ESL curriculum in senior secondary schools in Botswana were interviewed. While on the one hand Curriculum Development and Evaluation (CD & E) officers and Senior Education Officers (SEOs) generally regard English speaking in the Botswana linguistic situation to be just as important as reading or writing, Examinations, Research and Testing Division (ERTD) officers did not think that speaking in English was that important.

Methodology

It is a descriptive survey because content analysis, which is the key data collection method for the study, can be seen as a method for descriptive survey for which the subjects are documents whose characteristics are going to be described quantitatively. English Syllabi and English exam papers from 2000 to 2005 across Paper 1 to Paper 3 were collected from the randomly sampled schools for content analysis of the subject matter and cognitive skills called for in these documents. For English, 6 paper 1 and 6 paper 2 final examination papers and mid-year examination papers respectively were collected between 2000 and 2005. There were 6 mid-year examinations paper 1 and 6 mid-year examinations paper 2 collected for English.

The mark awarded to the item was translated as the mark awarded to the skill being measured and then recorded. Marks awarded to similar categories were added up for the entire test to make 13 numerical values representing each of Bloom cognitive subcategories. Human cognitive levels as defined by Bloom *et al*. (1956) rather than the terms themselves were used to operationalize cognitive behaviour to avoid invalid meaning of behavioural terms. Codes derived from the syllabus and from the examination papers were correlated using Pearson's r. Similarly, content codes in the examinations, subtopics addressed in the syllabi were extracted. Each subtopic was assigned a figure that showed how many times it is mentioned in the syllabus. Examination papers were analysed item by item and subtopics addressed in the examination items were paired with the mark awarded for the item in question. Codes resulting from the syllabus and those from the examination papers were then correlated using Pearson's r. Correlating codes from various examination papers with every other exam were further analysed. Unprocessed data analysis was necessary to appropriately take out the sequence of information with fewer mistakes. All figures analysed using SPSS.

Conducting training with three colleagues ensured rater reliability. A sample set of examination papers was selected and the three raters were asked to code them for skill measurement. Each rater coded a total of 72 items. Since the data derived was ordinal, inter rater reliability was calculated using Kedall's Tau, the Cronback alpha for the instrument when first person singular was used was found to be .948, but when the first person plural was used it comes to .860. The latter was disregarded. Descriptive statistics indicated that deviation in the coding was lowest. Again, all figures were analysed using SPSS.

Presentation of results

Analysis of Content of English called for by the Syllabus and Measured by Examinations in Botswana Senior Secondary Schools

Ho_1: The validity of measuring English content by midyear and by BGCSE examinations with the content of the syllabus, as the criterion is not significant.

The validity of measuring English content by midyear and by BGCSE examinations with the content of the syllabus as the criterion was significant in Botswana senior secondary schools from 2000 to 2005. The Pearson product moment correlation coefficient shows a high correlation (r =. 749 to .781, p <. 05) between the measurement of English content called for by the syllabus and measured by BGCSE examinations for year 2000 to 2005 (see Table 2). There is a low correlation (r = .473, p <.

05) between the measurement of English content called for by the syllabus and measured by midyear 2000 examinations.

Ho_2: In the measurement of content, the validity of English by mid-year examinations given the English syllabus as the criterion is not significantly higher than the validity of the English mid-year examinations given the English BGCSE examinations as the criterion.

To test this hypothesis, a Z – test of comparing Fisher's transformed r – values from dependent sample was done.

Table 2: Correlations in Content between the English Syllabus and English Mid-year and BGCSE Examinations (n = 336)

		cotent sylbus	Mid yr '00	Mid yr '01	Mid yr '02	Mid yr '03	Mid yr '04	Mid yr '05	bgcse 2000	bgcse 2001	bgcse 2002	bgcse 2003	bgcse 2004	bgcse 2005
cotent sylbus	Pearson r	1	.473	.569	.623	.681	.703	.745	.749	.781	.767	.771	.773	.754
	Sig.		.002	.005	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Mid yr '00	Pearson r	.473	1	.294	.206	.648	.646	.981	.996	.385	.993	.992	.992	.989
	Sig.	.002		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Mid yr '01	Pearson r	.569	.294	1	.212	.388	.267	.336	.334	.806	.343	.348	.348	.356
	Sig.	.005	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Mid yr '02	Pearson r	.623	.206	.212	1	.155	.007	.115	.192	.385	.197	.197	.198	.191
	Sig.	.000	.000	.000		.004	.904	.035	.000	.000	.000	.000	.000	.000
Mid yr '03	Pearson r	.681	.648	.388	.155	1	.381	.619	.625	.271	.626	.627	.627	.622
	Sig.	.000	.000	.000	.004		.000	.000	.000	.000	.000	.000	.000	.000
Mid yr '04	Pearson r	.703	.646	.267	.007	.381	1	.703	.667	.246	.668	.667	.668	.664
	Sig.	.000	.000	.000	.904	.000		.000	.000	.000	.000	.000	.000	.000
Mid yr '05	Pearson r	.745	.981	.336	.115	.619	.703	1	.988	.371	.988	.988	.988	.989
	Sig.	.000	.000	.000	.035	.000	.000		.000	.000	.000	.000	.000	.000

The results as presented in table 3 in measurement of content shows that the validity of English by midyear 2002 examinations, given the English syllabus as the criterion, was significantly higher ($Z_{.05}$ > 1.96, 6.8363) than the validity of the English mid-year examinations, given the English BGCSE examinations as the criterion. However, the validity of English by mid-year examinations, given the English syllabus as a criterion is not significantly higher ($Z_{.05}$ < 1.96, -32.889 to 1.2207) than the validity of the English

mid-year examinations, given the English BGCSE as the criterion for years 2000, 2001, 2003, 2004 and 2005 (see Table 3).

Table 3: Z-test Analysis of the Differences in the Coefficients of Pearson Correlation in Content between SMCMMYE & SMCPS and between SMCMMYE & SMCMBGCSE in English for 2000 – 2005

Year	Variables	r-value	Z_r (Z-transformed) value	Z-test value
2000	SMCMMYE & SMCPS	.473	.5142	-32.889
	SMCMMYE & SMCMBGCSE	.996	3.063	
2001	SMCMMYE & SMCPS	.569	.6464	-6.0569
	SMCMMYE & SMCMBGCSE	.806	1.1158	
2002	SMCMMYE & SMCPS	.623	.7298	6.8363
	SMCMMYE & SMCMBGCSE	.197	.2000	
2003	SMCMMYE & SMCPS	.681	.8308	1.2207
	SMCMMYE & SMCMBGCSE	.627	.7362	
2004	SMCMMYE & SMCPS	.703	.8730	.8465
	SMCMMYE & SMCMBGCSE	.668	.8074	
2005	SMCMMYE & SMCPS	.745	.9620	-21.2160
	SMCMMYE & SMCMBGCSE	.989	2.6062	

Analysis of the English Cognitive Skills called for by the Syllabus and Measured by Examinations in Botswana Senior Secondary Schools

Ho_3: Given the English syllabus as the criterion, the validity of midyear and BGCSE examinations in terms of the measurement of cognitive skill is not significant.

The validity of measuring English cognitive skills by midyear and by BGCSE examinations given the cognitive skills demanded by the English syllabus as the criterion was significant in Botswana senior secondary schools from year 2000 to year 2005

The Pearson product moment correlation coefficient shows a weak but significant correlation ($r = .379$, $p < .05$) between the English cognitive skills called for by the syllabus and measured by BGCSE 2001 examinations (See Table 4). On the other hand, there is a moderate correlation ($r = .634$ to $.685$, $p < .05$) between the English

cognitive skills called for by the syllabus and measured by BGCSE 2001 examinations. There is also a weak but significant correlation (r =. 381 to .427, p <. 05) between the English cognitive skills called for by the syllabus and measured by mid-year examinations for year 2001, 2002 and 2004.

Ho_4: In the measurement of cognitive skill, the validity of English by mid-year examinations given the English syllabus as the criterion is not significantly higher than the validity of the English mid-year examinations given the English BGCSE as the criterion:

Table 4: Correlations in Skill between the English Syllabus and English Midyear and BGCSE Examinations (n = 52)

		Skill sylbus	Mid yr '00	Mid yr '01	Mid yr '02	Mid yr '03	Mid yr '04	Mid yr '05	bgcse 2000	Bgcse 2001	bgcse 2002	bgcse 2003	bgcse 2004	bgcse 2005
Skill sylbus	Pearson r	1	.627	.396	.427	.532	.381	.630	.653	.379	.649	.636	.634	.685
	Sig.		.000	.004	.002	.000	.005	.000	.000	.006	.000	.000	.000	.000
Mid yr '00	Pearson r	.627	1	.624	.665	.827	.436	.959	.989	.120	.990	.957	.975	.984
	Sig.	.000		.000	.000	.000	.001	.000	.000	.398	.000	.000	.000	.000
Mid yr '01	Pearson r	.396	.624	1	.377	.622	.543	.658	.647	.028	.648	.625	.640	.644
	Sig.	.004	.000		.006	.000	.000	.000	.000	.843	.000	.000	.000	.000
Mid yr '02	Pearson r	.427	.665	.377	1	.841	.024	.610	.657	.118	.657	.637	.646	.654
	Sig.	.002	.000	.006		.000	.866	.000	.000	.406	.000	.000	.000	.000
Mid yr '03	Pearson r	.532	.827	.622	.841	1	.476	.781	.817	.157	.817	.807	.792	.814
	Sig.	.000	.000	.000	.000		.000	.000	.000	.267	.000	.000	.000	.000
Mid yr '04	Pearson r	.381	.436	.543	.024	.476	1	.440	.471	.359	.468	.489	.432	.471
	Sig.	.005	.001	.000	.866	.000		.001	.000	.009	.000	.000	.001	.000
Mid yr '05	Pearson r	.630	.959	.658	.610	.781	.440	1	.971	.042	.973	.941	.960	.976
	Sig. r	.000	.000	.000	.000	.000	.001		.000	.769	.000	.000	.000	.000

To test this hypothesis, a Z – test of comparing Fisher's transformed r – values from dependent sample was carried out. The results as presented in table 4 in the measurement of cognitive skill indicated that the validity of English by mid-year examinations given the English syllabus as the criterion is not significantly higher ($Z_{.05}$ <

1.96, -9.3550 to 1.9363) than the validity of the English mid-year examinations given the English BGCSE as the criterion in Botswana senior secondary schools (see Table 5).

Table 5: Z-test Analysis of the Differences in the Coefficients of Pearson Correlation of Cognitive Skill between CSBMMYE & CSBPS and Between CSBMMYE & CSBMBGCSE in English for 2000 – 2005

Year	Variables	r-value	Z_r (Z-transformed) value	Z-test value
2000	CSBMMYE & CSBPS	.627	.73620	-9.3550
	CSBMMYE & CSBMBGCSE	.989	2.6062	
2001	CSBMMYE & CSBPS	.396	.4192	1.9363
	CSBMMYE & CSBMBGCSE	.028	.0280	
2002	CSBMMYE & CSBPS	.427	.4564	-1.6394
	CSBMMYE & CSCMBGCSE	.657	.7876	
2003	CSBMMYE & CSBPS	.532	.5928	-2.6026
	CSBMMYE & CSBMBGCSE	.807	1.1186	
2004	CSBMMYE & CSBPS	.381	.4012	-.3029
	CSBMMYE & CSBMBGCSE	.432	.4624	
2005	CSBMMYE & CSBPS	.630	.7410	-7.2593
	CSBMMYE & CSBMBGCSE	.976	2.2076	

Discussion

Generally, English as a subject in Botswana senior secondary schools is one of those that show that teaching to the test maybe prevalent than thought. This is immensely because some content and skills taught to the students as per the syllabi is not examined. The validity of measuring English content and cognitive skills by midyear and BGCSE examinations given the English syllabus as the criterion is significant in Botswana senior secondary schools from year 2000 to year 2005. The significance of the validity in measuring content and cognitive skills by midyear and BGCSE examinations given the syllabus as the criterion suggest that teachers may be poor in item construction. However, the validity of English by mid-year examinations given the English syllabus as the criterion was not significantly higher than the validity of the English mid-year examinations given the English BGCSE as the criterion for year 2000, 2001, 2003, 2004

and 2005. The weak but significant correlation between the English cognitive skills called for by the syllabus and measured by mid-year examinations for year 2001, 2002 and 2004 shows that there was poor item construction of the examination since it was not representative of the syllabus.

This practice has shifted the focus of mid-year examinations to the extent whereby people view them as being nothing more than just an examination preparedness exercise which does not have to measure, as it rightly should, the level to which the syllabus has been covered before students sit for final examinations.

Speaking is a survival skill worldwide and yet it is one of the components not examined but taught in schools. The belief is that the skills taught to students is done by means of a communicative approach, thus, the skills are not taught in isolation but that does not change the fact that some are not examined. The results may as well be reflecting findings in a report by The World Bank Group (2001) that Western European examination systems spread as the French, British and Dutch empires expanded in the 19th century. Syllabi and examination papers from the 'home country' were used, usually unchanged, in the colonies. As countries have gained independence over the past fifty or so years, they have taken control of their school examinations. However, the assessment methods, and in some cases the syllabi, have remained largely unchanged. That means examination items from past papers need to be changed from year to year or else the examinations will lose meaning over time.

School examinations are more public examination valid than syllabi valid and something has to be done before things get out of hand. A larger part of the content from the English syllabus is taught in schools but not tested either in school tests or public examinations, mostly from the listening and speaking contents. The argument that examinations do not measure skills most essential for development reduces the level of confidence we have for them to be used to select the students who are adequately equipped for life after school. This stands in contravention of the theory of construct validation, which proposes that testing should have social or value consequences. This further contravenes the argument by Nenty (1996) that test validity is actually the level of confidence with which an examinees' test score could be used to infer the ability under measurement possessed by the examine. This also confirms the argument by some scholars that scores from students who graduate from senior secondary schools may not be used to judge students adequacy in terms of skills imperative for the country development needs.

Conclusion

The content and skills in the tests represented more of the national examinations than the syllabus, an indication of teaching to the test rather than syllabus-based teaching and learning. Results revealed close similarity between mid-year examinations and final examinations. It was also found that public examinations in Botswana are a poor reflection of content and skill as outlined in the syllabus and that midyear exams are more aligned to BGCSE exams than to the syllabus, leading to the conclusion that testing and consequently, teaching in secondary schools is more test based than curriculum-based.

Recommendations

Borrowing from IRT and CTT theories, the ministry of education in Botswana should consider not only providing short courses to teachers on item construction but also invest in item bank development as part of long term improvement in the quality of teaching, learning and most importantly, assessment.

References

1. Akiyama, T. (2004). Introducing EFL speaking tests into Japanese senior high school entrance examination. Unpublished PhD thesis, University of Melbourne, Melbourne, Australia.
2. An, X., & Yung, Y. (2014). Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It. SAS Institute Inc. SAS364-2014
3. Baker, F. B. (2001). The basics of item response theory. University of Wisconsin. USA. 2nd Edition. ERIC clearinghouse on assessment and evaluation publications.
4. Bloom, B.S., Englhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (Eds.) (1956). *Taxonomy of educational objectives, the classification of educational goals, Handbook I: cognitive domain*. New York: Longmans.
5. Botswana Examination Council (2011), Summary of results. Retrieved from http://www.bec.co.bw/past-results/2011/2011_bgcse_book_version.pdf/view
   Cotton, K. (1993). Developing employability skills. Retrieved from http://www.nwrel.org/scpd/sirs/8/c015.html

6. Elder, C. (1997) What does test bias have to with fairness? Language testing. 261-277
7. Fetogang, B. E. (2015) Assessment Syllabus and the Validity of Agricultural Education. *Research Journal of Educational Sciences,* 3(5), 1-9, (2015).
8. International Reading Association (IRA), (2005). Improving the quality of literacy in the content areas: situational analysis of secondary level education in Botswana. Retrieved from http://unesco.unesco.org/images/0014/001422/142225e.pdf
9. Isaacs, S. (2007). ICT in education in Botswana: Survey of ICT and education in Africa: Botswana Country Report. Retrieved from www.infodev.org_BOTSWANA_ICTed_survey1.pdf
10. Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. Retrieved from http://pricetheory.uchicago.edu/levitt/Papers/JacobLevitt2003.pdf
11. Jones, B. D., Jones, G. M., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999) The impact of high-stakes testing on teachers and students. SAGE Publications, *Phi Delta International*, 12 (Nov 1999) 199-203
13. Juma, J. (2005). The high price of free primary education. Retrieved from http://irenkenya.org/page.php?instructions=page&page-id=44&nav-id=18
14. Kober, N. (2002). Teaching to the test: good, the bad and who is responsible. Test talk for leaders. ISSUE 1. Retrieved from www.cepdc.org/pubs/TestTalk/CEPTTestTalk3.html
15. Lee, R. (1990). Human resources development in Africa. In L. Nadler & Z. Nadler (Eds.) The handbook of human resource development (2nd ed.). (pp 11-1.47). New York; John Wiley & Sons
16. Nenty, H. J. (1996). Advances in test validation. In G. A. Badmus, & P. I. Odor (Eds.), Challenges of managing educational assessment in Nigeria (pp. 59-69). Kaduna, Nigeria: National Conference on Educational Assessment
17. Nkosana, L. (2008). Attitudinal obstacles to curriculum and assessment reform. *Language Teaching Research* 12, (2), 287–312, (2008)
18. Ntuane, B (2005). Re-introduction of school fees represents the beginning of a roll back of Botswana's welfare state. An abridged version of a speech he delivered at a prize giving ceremony of Marulamantsi junior Secondary School, on Wednesday 12 October 2005. The Daily News, p,B4.
19. Mnthali, N.S. (1995). *An evaluation of the teaching and examination of literature in English in Botswana.* Unpublished master of education dissertation, University of Botswana.

20. Sweing, E., Crisp, V., Ahmend, A., & Pollit, A. (2002) Tales of the expected: The influence of students' expectations on exam validity. Paper presented at the British Educational Research Association Conference. Exeter, September 2002.
21. The World Bank Group (2001). The nature of public examination. Public examination system. Retrieved from http://www1.worldbank.org/education/exams/nature.asp