Original Research

Analysis of the Alignment of Primary Education Completion English Examinations with the Textbooks in Oromia Regional State

Motuma Hirpasa

Department of English Language and Literature, College of Social Science and Humanities, Ambo University. P.O. Box: 19

Abstract	Article Information
Concerned with the high rate of students' repetition at grate eight in Primary Education Completion Examination (PECE) in Oromia Regional State, this study was conducted to analyse the alignment of PECE in Oromia Regional State (from 2006 to 2009 E.C) with the content and objectives of the textbooks of grade 5 to 8 in English language. To achieve this objective, document/content analysis was employed to gather the data from the textbooks and the examinations. Besides, to substantiate this tool, unstructured interview was held with five individuals, who were selected on the bases of available and snowball sampling techniques because they were constructing the examination during different times for Oromia Regional State Education Bureau. To analyse the data, multistep document analyses were employed. First, contents of the textbooks were thematically organised into six major themes: Listening, Speaking, Reading, Writing, Grammar and Vocabulary. Second, the contents of the examinations' questions were also organized thematically. Then, the frequencies of the content in each six themes of the examination and the textbooks were analysed using frequencies percentages, and the extent of alignment between the frequencies of the contents of the textbooks and examination has been determined using Siegel and Spearman Contingency Correlation Coefficient. The amount of time allotted for contents of the textbooks was also analysed against the number of questions in each theme of the examination using the same statistics to test the level of alignment between text books and examination questions. The result of the study showed that alignment between the textbooks and the examination content has been found to be relatively very low since the overall Correlation Coefficient between the two observations within the last four years is 0.08, which shows a problem of mismatch, disproportionate or distorted distribution between the content of the textbooks and the examination questions. The straight forward to the questions is to base on the curriculum material (textbook) to minimize negative wash back effects of the examination and as a result, the students could pass the examination.	 *Corresponding Author: Motuma Hirpasa E-mail: maldamercy@yahoo.com

INTRODUCTION

Examination has been defined by Oxford Dictionary of English (2005) as a formal test of person's knowledge or proficiency in a subject or skill (Brown, 2003; Munoz, 2013; Palacio, Gaviria, & Brown, 2016). These scholars also use examinations and tests interchangeably according to the specific context. According to Brown (2003), examinations/tests are method of measuring a student's ability, knowledge or performance in a given domain for the evaluation of student's performance for purposes of comparison or selection. Tests are

also tools for teachers to know how much the students have understood the lessons (Munoz, 2013). Moreover, tests provide a chance for learning to take place by requiring students to think about what they have learnt in a new way (Adair-Hauck, et. al, 2006). Furthermore, tests are to discover how far students have achieved the objectives of a course of study, to diagnose students' strengths and weaknesses, to identify what they know and what they do not know and to assist placement of students by identifying the stage or part of a teaching program most appropriate to their ability (Munoz, 2013).

Tests have also a great role in the improvement of curriculum by offering the necessary information about the implementation of the instructional material and the development of curriculum objectives (Palacio, Gaviria, & Brown, 2016). Language teaching and testing are intertwined as it helps students create positive attitude, competitive temperament and mastery of language because it helps teachers in raising morale, getting reflections, diagnosing error, identifying thrust areas; enhancing effectiveness, and knowing future course of action (Adair-Hauck, et. al, 2006). Language testing approaches are axioms or correlative assumptions which provide method (a set of testing style), design (norm and domain) and procedure (technique and administration) to follow Richards & Rodgers (2003). Tests can be given at classroom, school and national or regional level. Classroom tests are more of informal that every teacher takes at every step-in teaching whereas school tests and national or regional examinations are programmed or scheduled that administered at identified place and time.

A good examination is purposeful, valid, reliable, objective, comprehensive, practical, expected, instructive and useful. Similarly, many scholars, for example, Brown (2005) and

Adair-Hauck, et. al, (2006) support the fact that test preparation should give a considerable attention to these qualities (Palacio, Gaviria, & Brown, 2016). Therefore, it is a necessity rather than a choice to consider these qualities in test preparation. Besides, a language test should be designed with primary course objectives and should cover from all components of a course (Adair-Hauck, et. al, 2006). It is, of course, expected that all of these principal qualities contribute to the good quality of a given test or examinations. These principal qualities are termed as the alignment of tests with the textbooks (Palacio, Gaviria, & Brown, 2016).

Examination should be aligned to the contents and objectives of the course it is supposed to represents. According to Brown (2007) and Palacio, Gaviria and Brown (2016), a test is said to have alignment with the textbook from which it has been designed when its content comprises a reasonable representative sample of language skills, structures or language components with which it is supposed to cover (Linn, & Miller, 2005). In other words, a test should be designed based upon a careful analysis of the outline of the course and adequately represent each portion of the analysis of outline to measure the content of a particular course of study. Alignment goes beyond content validity that includes comprehensiveness and backwash effect in which its content constitutes a representative sample of the language skills and structures (Adair-Hauck, et. al, 2006). What should be tested in language? Normally, tests are supposed to test objectives of a course in a language. The language objectives are mainly organized in to six general categories: listening, speaking, reading, writing, grammar and vocabulary-word formation (Palacio, Gaviria, & Brown, 2016).

The alignment between language testing and language textbook is undeniable (Palacio,

Gaviria, & Brown, 2016; Motuma, 2014). In order to judge whether the test has alignment or not, test designers need a specification of the skills or structures that it is meant to cover (Brown, 2007). This indicates that a comparison of test specification and test content is the basis for judgments as to alignment. This judgment should be made by the people who are familiar with a production of language teaching and a test. To confirm the issues raised here above, this question should be considered. Did the examination setters use test specifications?

Although various abroad and local studies carried out on alignment of examinations in the past, there is no study held on the alignment of primary education completion English examination with the textbook. For example, Notar (2004) from abroad studied the factors influencing the validity and reliability of teacher made test in Kenya. Locally, Asmare (2008) and Tibebe (1992) studied about validity of test at different times and sites. The findings of all the studies show that the examinations have no or weak content validity. Alemu (1983) and Lalisa (2017) studied a related problem to the present study. Alemu (1983) tried to assess the content validity of grades Six and Eight English National examination during the old curriculum. Unlike Alemu (1983), Lalisa (2017) is the only one who tried to assess the content validity of primary school education completion using content analyses after the current Ethiopian educational policy was set in 1994. However, his study was not comprehensive in which it focused on content validity only.

The present study focuses on the alignment of examinations during the last four consecutive years (2006 to 2009) with the textbooks they were prepared from (grade 5-8). The term alignment is a more comprehensive than content validity because it includes not only the analyses of the degree of relationship between the examinations and the textbooks, but also the analysis of comprehensiveness, proportionality, amount of time allotted and backwash effects (Palacio, Gaviria, & Brown, 2016; Palacio, et.al., 2015). Moreover, the study includes the analyses of course objectives and the factors affecting the examination setters to align the examinations with the textbooks. Furthermore, five examination setters were included as the sources of the data. According to Palacio et.al (2015), test alignment is said to be comprehensive when it contains an appropriate proportion of items from all aspects of the language or material to be tested (Brown, 2005). The number of items needs to be in proportion to the importance of the language area being tested, or in proportion to the problems of the students (Brown, 2005).

The current Primary Education Completion Examination (PECE), which is an achievement test, is given at the end of grade 8. In Ethiopian context, PECE is a regional examination for grade eight students. National Examinations are given at grade 10 (secondary school completion) and at grade 12 (preparatory school completion or entrance examination for university). Therefore, this study is the first phase or the beginning of the three similar analyses of examinations for PECEE, secondary school completion and preparatory school completion. The present study is hoped to contribute some means or knowledge for examination constructors to align examination with the textbooks and syllabus, and it can show ways to syllabus designers and textbook writers to modify, supplement or change the contents of the textbooks and syllabus.

This study was therefore mainly limited to the analysis of the alignment of the contents of PECEE with the textbooks because teachers, educated students' parents, school administrators, other school communities and students usually express their complaints for

the failure examination constructors to keep and maintain the qualities of examinations. Prompted by the harmful effects of the mismatch between the examinations and the textbook, the researcher also began reading about the alignment of examinations with the textbooks since he has begun teaching the course 'Language Testing/Assessment' for both TEFL and Afan Oromo postgraduate students. Hence, the researcher has identified that the alignment of the current Primary Education Completion English Examination needs to be studied profoundly.

Statement of the Problem

The core of the problem is the dissatisfaction with the present quality of teaching and assessing English as a foreign language at all levels of schools in Ethiopia (Ambachew, 2003; Alemu, 2009; Teshome, 2001). Teachers at all level of schools usually complain each other for their students' poor English Language ability and, finally pass on the baton of blame to other immediate lower subordinate level of school teachers (Ambachew, 2003; Motuma, 2014, 2015 & Piper, 2010:8). Moreover, the results of the classroom teachers' assessment and PECEE are not consistent for most of the students. As the result, the teachers of the subject become frustrated to confirm whether grade 8 students are as effective as the course writers would expect them to be or not. As a result, the repetition rate of PECEE is becoming increasing from year to year in Oromia National Regional State.

During different training occasions, second cycle primary school English language teacher usually complain that the repetition rate of primary education completion examination of PECEE is increasing from time to time for the students' relatively lower results in English language examinations highly contributed for the students' failure in PECEE. Besides, more

recently, the teachers have been complaining that the declining rate of the students' result in English language become worse and worse. Furthermore, according to teachers' comment, the result of most (80%) of the students in English language examinations is declining from year to year and from grade five to eight. Besides, the examinations included some language components but not in the textbooks by which the students were tested by the lesson they had not learnt.

Several local studies confirmed that the repetition rate in an examination can be increasing from time to time as result of the problem of alignment between the contents of the examination and the textbook (Samuel, 2002; Dejene, 1990; Teshome, 1995 & Kifle, 1995). These scholars justified that students' lower results in English language examinations can contribute for the higher rate of repletion in any examination. For the examinations to play their great roles on teaching and learning, they should possess different principal qualities of a good tests mainly high degree of alignment with the textbooks. Examinations with poor alignment with the textbooks have a harmful backwash effect on the entire process of teaching and learning. In other words, the consequences of mismatch between examinations and textbook are double: students could pay attention to only the content areas addressed in the examinations during the classroom instruction, and the big share of the preparation for the examinations will be restricted to those areas that frequently appear in the test. For this reason, this study has addressed the following research questions.

1. To what extent do the four textbooks of grades five, six, seven and eight and PECEE of 2006, 2007, 2008, and 2009 E.C inform each other?

2. What is (are) the effect(s) of the examinations on the teaching-learning process using respective textbooks?
3. What are the factors that hinder the examination setters to align the language elements of PECEE with textbooks of grades 5 to 8?

Research objective

The general objective of the study was to analyse the alignment of the contents of primary education completion English examination from 2006 to 2009 E.C with the contents of English textbooks for grades five to eight.

The specific objectives of the study were to:

1. Determine whether the 2006 to 2009 E.C examinations samples of language elements and all textbooks of grades 5 to 8 inform each other.
2. Analyse the effects of examinations on teaching-learning process using the respective textbooks.
3. Identify the factors that affect the examinations setters to align the examinations with all the language elements of all the textbooks of grades 5-8.

Delimitation and Limitation of the study

The present study is confined to the analysis of the alignment of the primary education completion English examinations within four consecutive academic years (from 2006 to 2009 E.C) with the textbooks of grade five to eight. This study has also highlighted the implication of the mismatch between examinations and textbooks (wash back effect) and factors affecting the examination setters to include the whole contents of the textbooks in the examinations. However, this study has not included item analyses: items level of difficulty and power of discrimination of the last four years back from now.

Research Design and Methodology of the Study

In the view of pragmatist research paradigm, a mixed research design was employed to describe the alignment of the examinations with the textbooks. Document analyses research method was utilised to gather qualitative and quantitative data from the subjects of the study (Nevendor, 2017). The main sources of the data were documents: English textbooks from grades 5 to 8, the PECEE of 2008 to 2009 E.C. and five English examination setters for Oromia Regional Education Bureau. All the sources of the data were selected for the study on the basis of purposive sampling technique because they are the basic sources to reveal the alignment of the examination with the textbooks and the factors that may hinder the examination setters to make the examinations comprehensive (Johnson & Turner, 2007).

Two instruments were used to collect both qualitative and quantitative data from the sources of the data. Content analyses (examinations and textbooks analyses) were the major tool to collect qualitative and quantitative data that have shown the alignment of the examinations and the textbooks. This study employed *Integrative Model* of Content Analysis as a framework of the study because it links data to logic from the examinations to the textbooks (Nevendor, 2017). Hence, two content analyses were made to identify to what extent the contents of the examinations were aligned with the textbooks: 1) analysis of grade eight examinations given within 2006 -2009 E.C in Oromia Regional State and, 2) analysis of textbooks for grades 5 to 8 (Nevendor, 2017). In addition to these analyses, unstructured interview was used with five individuals who constructed grade eight regional examinations to collect data on the factors that affect the

examination setters to align all the examinations with all textbooks. This is because interview is powerful way and helpful instrument to collect in-depth and flexible qualitative data (Dornyei, 2007; Shohamy, 2004).

As the data collecting and analysis procedures, the contents in all English textbooks (Grades 5, 6, 7 and 9) were first separately reviewed (book review) and thematically organized in to six language components: listening, speaking, reading, writing, vocabulary and grammar. Book review has helped the research get a brief summary of the contents of Grades 5, 6, 7 and 8 English textbooks (Nevendor, 2017). Based on the themes of contents identified in each textbook, the contents of primary education completion English examinations, given within the year 2006 - 2009 E.C, were also separately organized in to six language components according to the themes of the contents of the textbooks (Sarantakos, 2005). The alignment of between the objectives of the textbooks and the examinations mainly analysed using protocol analyses of the coverage, representativeness and correlation between the items of every examination with the contents and the time allotted to every theme of every textbook (Nevendor, 2017).

Then, the summary of the frequencies of the language content in both textbooks and examinations were presented in figures and percentages in separate tables. For instance, dialogue completion is considered as one contents of speaking in both textbooks and examinations. Therefore, if there might be 5 dialogue exercise The formula is:

questions, the research counted it as 5 frequencies of the speaking in textbook. In the same token, five dialogue questions in an examination represent five frequency of the speaking section of the examination. As a result, the percentage of each frequency was separately calculated for the textbooks and the examinations. The comparisons of these percentages of the examinations against the textbooks and different types of graphs were used to show the extent of alignment, comprehensives and proportional coverage between the frequencies of the contents in the textbooks and frequencies of the contents of the examinations (Nevendor, 2017). Similarly, the percentage of time allotted to each language components were used to analyse the alignment of examinations with the textbooks.

Mainly, to determine the extent of relationship/alignment between the frequencies of the contents in the textbooks and frequencies of the contents of the examination, Sidney Siegel's formula (contingency coefficient 'C') was also used to determine the degree of alignment between the textbooks and the examinations. Therefore, applying Sidney Siegel's formula, the next procedure is putting these two sets of the data (i.e., contents of textbooks and the examinations) in to different columns of a table to compare them to each other (Nevendor, 2017). This helps to determine the degree of relationship or alignment between the two sets of data (Harper, 1977, cited in Palacio, Gaviria, & Brown, 2016) and Sarantakos (2005).

$$C = \sqrt{\frac{X^2}{N+X^2}}$$

Where, C = the value of relationship,

N = grand total,

X^2 = refers to Chi-square- determined after employing the following three steps,

Step 1: Preparing contingency table (e.g. see Table 2, sets of observations of six (6) components: reading comprehension, vocabulary, Grammar, speaking, listening and writing.

Step 2: Determine "expected value" (see Table 2): is computed using the following formula:

$$\text{Expected value } (E) = \frac{Rowtotal(r) x (c)Colomntotal}{Grandtotal (N)}$$

In step3: Determine value of Chi-square (X^2): The formula is:

$$x^2 = \frac{(observedvalue (O) - (E)expectedvalue)^2}{epectedvalue(E)}$$

Table 1: The Interpretation of the relationships between two observations

S/N	Value of contingency coefficient	Interpretations	Effects
1	C = 1	A perfect correlation, comprehensiveness	Positive
2	0.91 -0.99	very high correlation, comprehensiveness	Positive
3	0.71 - 0.91	High correlation, comprehensiveness	Positive
2	C = 0. 41 to 0. 70	Moderate correlation, comprehensives	Positive
3	C = 0.21 to 0. 40	Low correlation and comprehensives	Negative
4	C = > 0. 2	Very low correlation and comprehensives	Negative

RESULTS AND DISCUSSION
Textbook Analysis

English textbooks for grades 5 to 8 have been published first in 2011 and reprinted in 2012. These textbooks have 12, 13, 12 and 14 units respectively which of each unit of the four textbooks have been divided into different lessons. The major language content areas of these textbooks are reading comprehension, vocabulary practice, grammar/language pattern (sentence structure and forms), speaking (dialogues), listening (riddle comprehension, gaps fill-in activities) and writing (rearranging jumbled) (English syllabuses and Textbooks, 2011). The sequence of the contents of the textbooks varies from textbook to textbook and from unit to unit within a textbook. The focus of this study was English language examinations, given in the year 2006 to 2009 E.C, were the second, third, fourth and fifth examinations respectively since the publication of the textbooks in 2011. This is because of the current second cycle primary school education includes all textbooks from grade 5 to8 from which the questions of the examinations were set(Table 1).

It was difficult to put exactly the time for the theme of the contents separately in terms of frequencies because the practice of some of the items of some the contents/skills are either too general or overlapping with each other. However, attempts were made to count the contents and the periods allotted to them as closer as possible to the reality and categorize them based on the judgement of the researcher. The frequency of the time given for grammar exercises for the students' practice in the textbooks, for example, is 170 from 613 periods to all the four textbooks. The time given for grammar covers 27% of the time allotted to the six major contents mentioned elsewhere. The detail information that reveals the total number of periods allotted to the total frequencies of contents in the six themes of the textbooks for grade 5-8 is indicated below in Figure 1
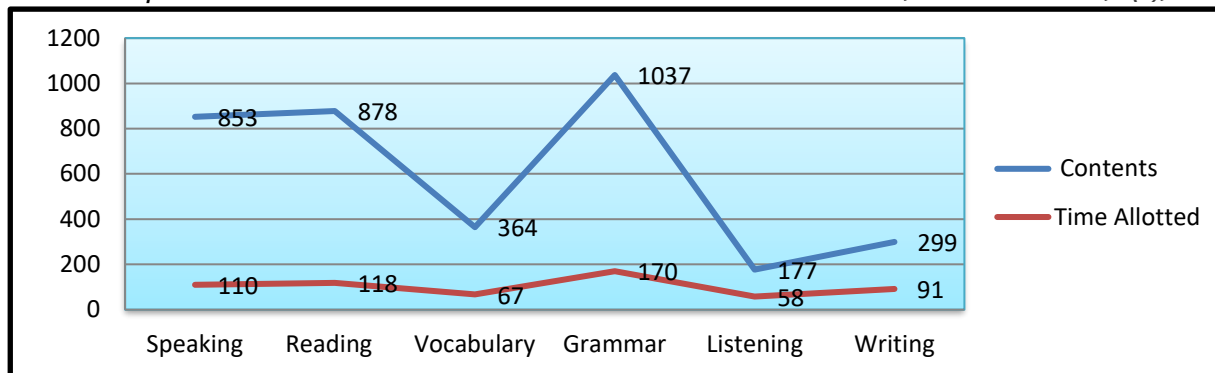
.Figure 1: Total Number of Periods allotted to the Total frequencies of Contents in the Six Themes of the Textbooks for Grade 5-8

The difference in emphasis among the language content areas becomes more colorful when they were compared and contrasted with the amount of time given to them. The amount of the time allotted to the contents of the textbooks seems to be almost proportional to content. Figure 2 shows that grammar 170 (27%), reading 118 (19%), speaking 110 (18%) were given from the total of 613 periods. However, much less emphasis is given to writing 177 (15%), vocabulary 67 (11) and listening 58 (10%) respectively in descending order. Hence, the following Figure 2 below shows the analysis of the proportion of the time allotted to the six major language content areas.
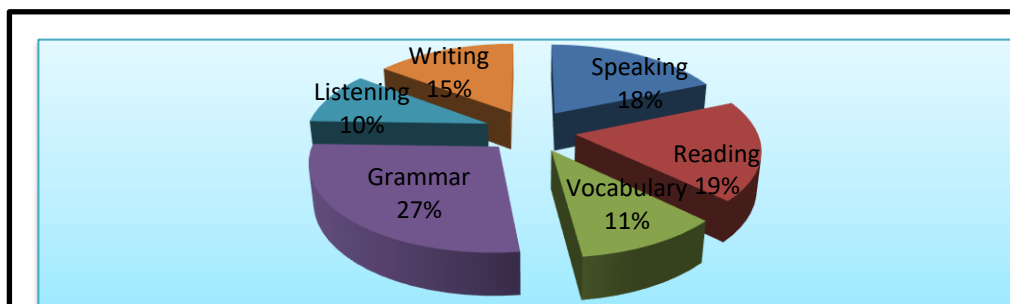

Figure 2: The percentage of Total Periods allotted to the six major Language Components areas.

In relation to the analysis of the frequency percentages of the time allotted to each of the six major language content areas of the four textbooks as it is indicated in Figure 2, the results show that much more emphasis was given to grammar over the language skills: reading, speaking, writing and listening. This indicates that the curriculum design and development have been based on the structural approach in which language skills are relatively given less emphasis. The next, section presents the analyses of PECEE given during 2006 to 2009 E.C.

Analyses of the Examinations

The analyses of contents in the four textbooks for grade 5 to 8 are the bases for the analysis of contents in the examination questions because the contents of each item of the examinations need to be classified according to the major language content areas in the textbooks. To achieve the representativeness of the examinations, the

researcher counted all the items of the examinations given during 2006 to 2009 E. C. and thereby related them to the percentages of the time given to each theme of the contents in the textbooks. This is because the researcher believes that the number of periods given to language contents can indicate the emphasis given to a particular content area. From the contents of the English examinations (2006 to 2009 E.C), a total number of 60 items were set each year and a total of 240 items for the last four years. Hence, the following Figure 3 shows the analysis of comprehensiveness and proportionality of the items of the examinations (2006-2009 E.C.) to the six major language content areas in terms of periods allotted to them.
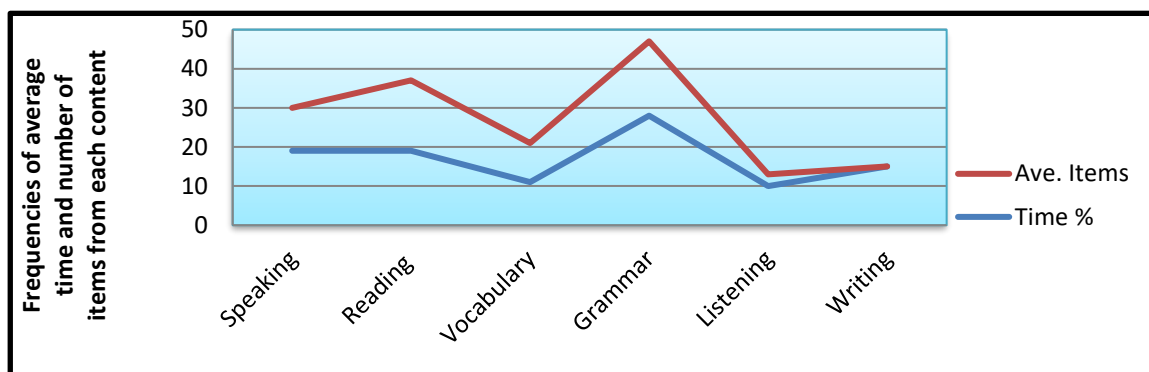

Figure 3: Total periods allotted to the six themes of language content and the items of Examinations

Figure 3 shows the frequency percentages of the items of the examinations and the frequency percentages of the time given to the students to learn the skills in the textbooks were respectively grammar (38.33%, 27.7), reading (25.42%, 19.25%), speaking (17.08%, 17.95%), vocabulary (16.25%, 10.93%), writing (2.92%, 14.85%) and listening (0%, 9.46%) in descending order. More importantly, the representativeness of the examinations to each sample of skills and content areas of the instructional materials shows a serious problem. For example, listening skill was totally ignored in the examinations because, according to the data obtained through the interview, it is difficult to test listening skills indirectly. The next step was to determine the degree of relationship between the classifications of the content areas of the textbooks with English examinations.

Relationships between Textbooks and Examinations
There were some mismatch representation and disproportion between the textbooks and examinations with regard to representativeness of these contents. In the textbooks, grammar, reading and speaking, constituted respectively in descending order higher coverage. In general, three sets of data in terms of frequencies have been presented in Figure 4 using figures: data from the analysis of the textbooks (grade 5-8), the data analysed from the examinations of the years from 2005 to 2008 E.C and the data analysed from the time allotted to cover the textbooks.

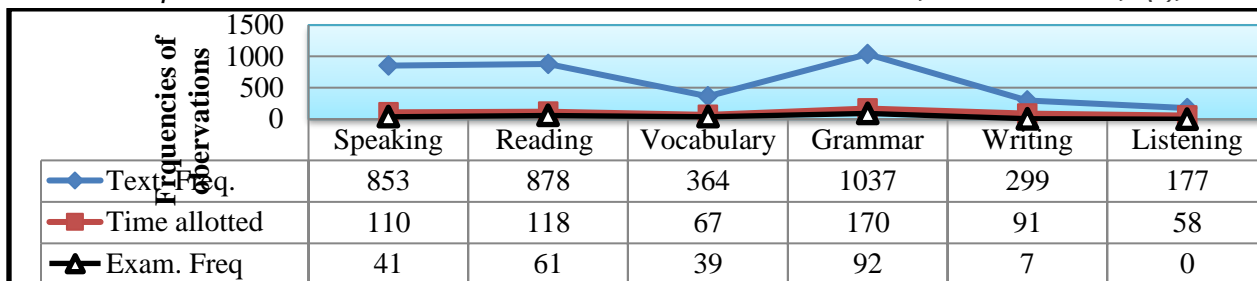|		Speaking	Reading	Vocabulary	Grammar	Writing	Listening
◆	Text Freq.	853	878	364	1037	299	177
■	Time allotted	110	118	67	170	91	58
△	Exam. Freq	41	61	39	92	7	0

Figure 4: The relationships between the textbooks' language content, periods allotted and the items in the examinations

Figure 4 shows the comprehensiveness and proportion between the frequencies of all examinations' items and the total frequencies of items of practice in the four textbooks. Cumulative comprehensiveness of the primary education completion English examinations for the last four consecutive years (2006 -2009 E.C) is calculated to compare it with the total frequencies of content areas of grades 5, 6, 7 and 8 English textbooks. Cumulative comprehensiveness shows the degree of alignment between the textbooks and the examinations. As indicated in methodology section, alignment is calculated for the frequencies of the textbooks and the frequencies of examination for each specific theme of the contents, as well as for each grade and year. Once the total frequency of the textbooks is calculated, the result (3668) is always the same and constant for all the years and grades. The number of the items of the examinations (60) is also the same for each year (2006 to 2009 E.C.). However, the number of activities and amount of time allotted to each language contents within the textbooks and the number of items within the examinations varies from grade to grade and from year to year.

The next procedure is putting these two sets of data (contents of textbooks and the examinations) which were selected to be compared to each other in a table 2 to determine the degree of extent of alignment between the two sets of data separately by applying Sidney Siegel's formula (see methodology section). The frequency of the major contents in the four textbooks and examinations for the last four years are summarized in table 2 which indicates the highest coverage of grammar throughout the textbooks and the examinations. In the following tables, the following abbreviations are used. TbOV means textbooks observed value; EOV means examinations observed value; TOV means total observed value; TEV is to mean textbooks expected value; EEV is used for examination expected value.

Table 2: Summary of the Major Content of the Examination and the Textbooks of Grade 5-8 (Summary of Contingency Tables)

NO	Contents	TbOV	EOV for each year				TbEOV	EEV for each year			
			2006	2007	2008	2009		2006	2007	2008	2009
1	Speaking	853	11	10	10	10	863	14.1	14.1	14.1	14.1
2	Reading	678	20	11	20	10	898	11.5	11.3	11.4	11.3
3	Vocabulary	364	10	9	10	10	374	6.1	6.1	6.1	6.1
4	Grammar	1037	19	28	20	25	1057	20.5	20.7	20. 6	20.6
5	Writing	299	-	2	-	5	299	4.9	4.9	4.9	4.9
6	Listening	177	-	-	-	-	177	3.0	4.9	2.9	2.9
	Total	3608	60	60	60	60	3668				

The data in Table 2 presents the sum total of the two frequencies, which gives the row total (total frequency of examinations, 60 for each) and column total (total frequency of the textbooks, 3608 constants for each year. The next step is calculating the correlation of

examinations have with the contents of the textbooks using the specific information in every year and grade. The table shows that grammar had the highest number of items both in examinations and in the textbooks, particularly in 2007. However, writing has the least and imbalanced coverage (7 items out of 240) but 299 activities in the textbooks, whereas there is no listening question throughout the examinations. These indicate that there was clear disproportion among the content areas of the textbooks and examinations in general. Table 3 also shows the analyses of both the four textbooks from grade 5 to 8 and the examinations given during 2006 to 2009 E.C.

Table 3: The Relationships between the Contents of the Textbooks of grade 5-8 and the Examinations in 2006 2008 E.C. by Sidney Siegel's Contingency Coefficient

S/N	Years in E.C.	Relationship By Grades			Remarks
		N	X^2	C	
1	2006	3668	17.195	0.069	
2	2007	3668	9.93	0.052	
3	2008	3668	18.24	0.07	
4	2009	3668	7.736	0.046	

The information in table 3 reveals the degree of the relationships between the contents of the four textbooks and every year examinations' items separately by year and grade. The degree of the extent of relationship between the contents of the four materials and the examination found to be 0.069, 0.052, 0.07 and 0.046, respectively for the years 2006, 2007, 2008 and 2009. Table 4, which is known as 'summary of contingency table', presents the overall summary of the textbooks and the examinations contents in general.

Table 4: The Overall Summary of the Textbooks and the Examinations Contents, Summary of Contingency Table

No	Contents	TOV	%	OVE	%	TOV	TEV	EEV
1	Speaking	853	23.6	41	17.1	894	848.88	14.117
2	Reading	878	18.8	61	25.4	939	877.73	11.27
3	Vocabulary	364	10.2	39	16.3	403	366.9	6.1
4	Grammar	1037	34.3	92	38.3	1129	1044.31	20.7
5	Writing	299	8.8	7	2.9	306	296.08	4.9
6	Listening	177	4.9	-	-	177	174.105	2.89
	Total	3608	100	240	100	3671		

Table 4 shows the conclusion of the whole analysis of the textbooks from grades 5 to 8 and the PECEE given from 2005 to 2008 E.C. The table has shown that there was a gap between the frequencies of contents of the textbooks and the examinations. Grammar as usual covers the highest positions in similar coverage. Whereas, in both textbooks and in the examinations reading comprehension covers the second positions preceding speaking in the examinations. Speaking and vocabulary shared the third and the fourth while writing covered the least. The overall summary of alignment between the four textbooks of 5-8 and all the examinations of 2006-2009 E.C. by Sidney Siegel's contingency coefficient indicated below.

$$C = \sqrt{\frac{X^2}{N+X^2}} \; ; \text{ where, } x^2 = \mathbf{23.856}$$

$$C = \sqrt{\frac{23.856}{3671+23.856}} = 0.08$$

The overall result of Sidney Siegel's contingency coefficient *(C = 0.08)* reveals that the alignment between the English

examinations given during 2006-2009 E.C and the English textbooks of grade 5-8 is very low.

Discussions

The alignment of the examinations and the textbooks depends on a careful analysis of the objectives and contents of language being tested. Thus, the relationship between the examinations' items and the textbooks' activities as well as the time allotted to achieve each objective and to cover each content areas in the textbook have been analysed. The alignment between language examination and language teaching is an obedient servant since it follows and apes teaching. To this end, an examination should have comprehensive covering and an appropriate proportion of items from all aspects of the language or material to be tested. The number of items needs to be in proportion to the objectives of the language area being tested, or in proportion to the problems of the students.

The separately computed results of contingency coefficient 'C' is 0.069, 0.052, 0.07, 0.042 and 0.08 for 2006, 2007, 2008 and 2009 E.C. respectively show that the degree of alignment between the contents of the textbooks and the examination is very low for all the examinations and the textbooks. However, the coverage and comprehensiveness vary from the themes to themes within and between the examinations and the textbooks. Grammar has dominated the items of examinations every year followed by reading sections. The results of the study show that grammar covered the highest number of items in both examinations and textbooks. In contrary to this, writing has 8.8% coverage in the textbooks but has the least and imbalanced coverage (2.9%) in examination, and listening was not included in to the examinations. Listening in textbooks coverage contains 4.9% but 0% in all the last four years examinations. The result stated in Figure 1 to

4 shows that except for the speaking skills (17.08%, 17.95%), there is mismatch between the items of the examinations and the time allotted contents in the textbooks.

The comparison between the coverage of skills and time allotted to cover the contents was also analysed to estimate the alignment. The result stated in bracket next to each language skills compares the coverage of skills and time allotted to it. Hence, grammar (38.33%, 27.7), reading (25.42%, 19.25%), speaking (17.08%, 17.95%), vocabulary (16.25%, 10.93%), writing (2.92%, 14.85%) and listening (0%, 9.46%) in descending order cover the textbook. The comparison shows that except speaking skills (17.08%, 17.95%), mismatch between the contents and the time allotted to learn the contents are apparent. What do these discrepancies between the items and the time allotted for the contents tell us? These may tell us two things: the examinations have given more unnecessary emphasis to some contents but undermined other contents both of which are indicating problems with the examinations to test whether the students have achieved the objectives stated in the textbooks or not. These in turn indicate that there was a clear disproportion among the content areas of the textbooks and examinations in general.

It has been stated elsewhere that a contingency coefficient value, C = 1, indicates a perfect correlation but less than 0.2 'C' values shows very low correlation between two observations. Therefore, the responses given to interview and questionnaire were computed and summarized. The summary of the responses indicated that there were factors that hinder the examination setters not to make the items proportional to the textbooks. Thus, the computed value of 'C' 0.069, 0.52, 0.07, 0.046 and 0.08 for each examination for all the years claim the same conclusions that the alignment between the two observation is very

low. All the respondents to the interview agree that unfair emphases given to grade 8 and 7 in a descending order.

Taking the textbook coverage as a constant factor, the contents of the examinations do not adequately represent the textbooks' contents. In other words, the examinations have limitation in describing and informing the outcomes of learning and teaching because the examinations fail to cover the broad learning outcomes of cognitive, affective and skills stated in the textbooks. In the same token, the examinations have limited opportunities for the candidates who learned using these textbooks to demonstrate more their skills and knowledge of the language. Thus, the under represent set up of the English examination, a true assessment of the extent of a candidate's knowledge of the language remains under question until the examinations were made to represent more of the textbook coverage. This is because the greater an examination's content comprehensive, the more likely it is to be an accurate measure of what it is supposed to measure; otherwise, the examination is unlikely to be accurate at all.

The summary of contingency table 4, which shows the analysis of contents of textbooks and the examinations both in figures and percentages, shows a clear disparity and disproportional allocation of the contents' percentages between the textbooks and examination as well as within the contents of the examinations. For instance, grammar constituted 28.7% of the textbooks and covers 38.3% of the examinations. Moreover, grammar and reading covered about 64% (25.4% + 38.3%) in the total examinations' items, but they cover 153(53%) in the textbooks. The responses given to the interview questions also justifies the complains with the incompatibility between the examinations' items format and contents to be assessed, like speaking, writing and listening which are difficult to test directly using multiple choice questions. As a result, 'dialogue completion' and 'rearranging jumbled sentences' were designed with the intention of assessing students' communicative ability. This indicates that the examinations were not constructed as to contain a representative sample of the textbooks throughout the years. There were also reciprocal relationships between speaking and reading with the degree of coverage between the textbooks and the examinations. Speaking and reading constituted 24.86% and 19.76% respectively in the textbook, however, they respectively constituted 17.08% and 25.4%. This is because, according to the respondents, reading comprehension questions are easy to set which helps one to test the students' ability to understand a gist of a paragraph and to extract key information on specific points in an examination. Similarly, writing and vocabulary constituted 10.08% and 8.71% in the textbooks' coverage respectively, but in the examination, they respectively covered 16.25% and 2.9% examinations. Writing ability is not easy to test in a form of closed item questions. This implies that there is inconsistence between the two materials in the areas of coverage. Moreover, it is also true that there were a lot of content areas given a bigger proportion in the textbooks but left unattended in the examinations. For example, listening skill has 177 (4.9%) items of practice and 58 (9.51%) periods allotted in the textbooks. Despite the coverage indicated, it was totally ignored from the total of 240 items in the examinations.

The overall alignment between the textbooks and the examinations were careful analyzed to identify the contents of the language being tested that may indicate the achievement of all textbooks' objectives. As a result, the representativeness and the relationships between the major themes of

contents of the four textbooks of grade 5 to 8 and the examinations given in 2006 to 2009 E.C. have been determined using percentages of the frequencies and time allotted as well as the contingency coefficient 'C' of Siegel and Spearman. The overall correlation between the textbooks and the examinations was found to be C = 0.08 which is judged to be very low because the value of the correlation coefficient was lessen 0. 2.

Several factors influenced the examination setters not to align the items of the examinations with the contents of the textbooks. The respondents for the interview confirmed that the examination constructor's educational background and total dependency on the previous table of specification and examinations were mostly affecting the exams setters. This is because, according to the respondents, the examination setters are enforced to use the contents of the preceding examinations' contents for the constraints of time and budget. The examination setters' role is only setting the examination, and the items were added to the exam bank. Therefore, a group of people who are selecting the items for the current year may fail to consider the existing situations of the examination takers supposed to take. Nevertheless, assessing candidates' speaking or writing ability through indirect testing (e.g. multiple-choice) format cannot be regarded as sufficient. However, every year, the examinations employed solely multiple-choice format which affected to include speaking, listening and writing skills. Particularly, listening speaking and skills are tested through auditory comprehensions in which short utterances, dialogue talk and lectures are given to the test taker. Peaking ability usually appears in the form of an interview, a picture description, roe play and problem solving task which involves pair work or group work.

This improper representation of items seems to negatively affect the achievement of beneficial backwash effect and results in redundancy of contents from year to year as well as the fulfilment of learning objectives and above all, lower the alignment of the examination with the textbooks. An academic intelligence regarding the subject area also has an impact on the examination setters to proportionate the contents of the examinations with the contents of the textbooks. Moreover, all of the interviewees said that uncertainty to cover all the four textbooks at all schools within a year negatively affected the representativeness of the examinations. In connection to this, the principle of keeping a balance in assigning weights to the areas to be tested is must if we need the information which results from performance of students on achieving examination as an indication of the degree to which the expected objectives of the course are being attained.

In summary, an attempt was made to summarize some of its major findings. The first major finding is related to determining the extent of relationship between contents of the grades 5, 6, 7 and 8 English textbooks and grade eight completion English examinations from 2006 to 2009 E.C of Oromia Regional State. Thorough analysis of each content area on the extent of relationship between the two observations has been found to be very low as it was interpreted in the finding and discussion sections. The second core of the finding related to the effects of the alignment of the contents of the textbooks and the examinations. The disproportional distribution of items or mismatch allocation of percentages to content areas in the examinations with the textbooks has negatively affected/negative backwash effect on the teaching-learning process. The third finding of the study was about the factors affecting the examination

setters to align the contents of the textbooks with the examination items.

Conclusions

Based on research objectives and major findings of the study discussed so far, language contents of the textbooks and the examinations did not inform each other. In other words, the areas coverage of contents in the examination and the allocation of percentages to content areas were characterized by disproportionate distribution that failed to adequately and proportionally represent the different content areas predominantly covered in the textbooks. The degree of the alignment between the textbooks coverage and examination content areas was found to be very low. The examinations did not possess the fundamental qualities required in good language examination of all kinds. This harmful mismatch between the PECEE and the textbooks had created negative backwash effect on teaching and learning processes. Relying on the previous examinations and the table of specifications usually leads the exam setter to give unfair emphases to some skills/textbooks. The incompatibility between the examinations' items format and contents to be assessed was resulted by students and teachers' academic intelligence, perceptions and designing skills of examinations, uncertainty to cover the textbooks within a year.

Embarking on the harmful effect of examinations construction, the examinations writer should first carefully draw up a table of test specifications and then, design the examinations describing in a very clear and precise terms to include the particular language objectives, skills and component areas in the examinations. It is advisable for Oromia Regional Education Bureau to make a content analysis scheme to determine alignment of every examination with the textbook and maintained the distribution of

percentages among content areas to be tested during the following examinations. Oromia Regional Education Bureau should give trainings for test constructors regarding language testing.

REFERENCES

Adair-Hauck, B., Glisan, E. W., Koda, K., Swender, E. B., & Sandrock, P. (2006). The Integrated Performance Assessment (ipa): Connecting assessment to instruction and learning. *Foreign Language Annals*, *39*(3), 359-382. http://dx.doi.org/10.1111/j.1944-9720.2006.tb02894. x.

Alemu Hailu. (2009). Ill-designed Reading Activities: The Case of a Unit in Grade 11 English Textbook, *The Ethiopian Journal of Education,24*(2), IER: AAU, Ethiopia, ISSN: 0425-4414.

Alemu Tsegaye (1983). *An Assessment of Grade Six and Eight English National Examinations* (MA Thesis). Addis Ababa University.

Ambachew M. D. (2003). *A Review of the Ethiopian Related Literature on Reading Ability, 83-93.* University of Pretoria.

Asmare Mehret (2008). *An Assessment of the Content Validity of English Language Tests: The Case of Awassa Health Science College* (MA Thesis). Addis Ababa University.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment.* New York, ny: McGraw-Hill.

Brown, J. D. (2012). *Choosing the right type of assessment.* In C. Coombe, S. J. Stoynoff, P. Davidson, & B. O'Sullivan (Eds.), *The Cambridge guide to language assessment* (pp. 133-139). Cambridge, UK: Cambridge University Press.

Brown.H.D.(2003) *Teaching by principles " an interactive approach to language pedagogy."*

C.R. Kothari (2006) *Research methodology /methods and techniques*. University of Rajasthan, India

Dejenie Leta (1990). *Achievement, Wash back and Proficiency in School Leaving Examination:* A Case of Innovation in an Ethiopian Setting. (Unpublished PhD, Thesis). Lancaster: Lancaster University

Devidson, F. (2002) *Test craft: a teacher's guid to writing and using language test specifications New Heven:* CT: Yale University press.

Dornyei, Z. (2007). *Research methodology in applied linguistics. Quantitative, qualitative and mixed methods.* Oxford: oxford university press.

Kifle Kebede (1995). *An Assessment of Content Validity of High School English Language Tests in Relation to Textbooks Concerned* (MA Thesis).Addis Ababa University.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, nj: Prentice Hall.

Motuma, H. (2014). An assessment of the Students' Reading Ability in Higher Institutions: The Case of Asela College of Teacher Education. *Journal of Science and Sustainable Development 1*(2), 76-95.

Motuma, H.(2015).The Changing Role of the University TEFL Teacher in Improving the Students Communicative English Skill: From Passive Technician to Reflective Practitioner. *Journal of Science and Sustainable Development*, *3*(1), 133-155.

Muñoz, A. (2013). *Beyond testing. Unpublished manuscript, Centro de Idiomas.* Universidad eafit, Medellín, Colombia

Nevendorf, A. K. (2017). *The content Analysis Guidebook* (2nd Edition), SADE Publishing Cleveland State University, PP 138.

Notar, C.E. (2004). *The table of specification: insuring accountability in teacher- made test. Journal of instructional psychology*, june.

Oxford Dictionary of English (2005) *The World's Most Trusted Dictionaries (2nded.).* Oxford: Oxford University Press.

Oxford Dictionary of English (2005). The World's Most Trusted Dictionaries (2nd ed.).Oxford: Oxford University Press.

Palacio, M., Gaviria, S., & Brown, J. D. (2016). *Aligning English language testing with curriculum. profile Issues in Teachers' Professional Development*, *18*(2), 63-77.http://dx.doi.org/10.15446/ profile.v18n2.53302.

Palacio, M., Kashani, S., Gaviria, S., Salazar, R., Estrada, A., Botero, A. M, Brown, J. D. (2015). Report on the 2014- 2015 Idiomas-eafit English testing program. Medellín, co: Universidad eafit

Piper, B. (2010). Ethiopian Early Grade Reading Assessment, *Data Analysis Report: Language and Learning.* Ed Data II, Task order 7&9, USAID/Ethiopia.

Richards, R. & Rodgers, C. (2003) *Approaches and methods in language teaching.* Cambridge University Press.

Sarantakos. D. (2005). *Social Research(3rd ed.*).Online internet. Available: http://www. palgrave.com/ sociology/sarantakos

Sharma, R (2000) *Fundamentals of educational research*. Mcerut: Inter publishing House

Shohamy, F. (2004), *Reflections on research guideline, categories, and responsibility.*

Teshome Demisse (1995). *The Construction and Validation of Tests in English for Tertiary Education.* (PhD dissertation). Addis Ababa University.

Tibebe Alemayehu (1992). *The Predictive Validity of Ethiopian School Leaving Certificate English Examination and the Integrative Tests. Comparative Study*, *12*(1).

USAID-AED/ IQPEP, (2010). *Continuous Assessment and How to Use it: Teachers Module 3.AED*: A. A. Ethiopia.