

## **Content Validity of EFL Teacher-Made Assessment: The Case of Communicative English Skills Course at Ambo University**

**Motuma Hirpassa Minda\***

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

**Article History:** Received: August 9, 2018; Revised: October 2, 2019; Accepted: October 23, 2019

**Abstract:** Prompted by increased concerns about the effects of mismatch between assessments and the teaching materials, this study evaluated the content validity of teacher-made assessments of Communicative English Skills Course at Ambo University. To this end, three years (2016, 2017 and 2018) teacher-made assessments and teaching materials of the course were purposively selected. In addition to the documents, twelve English language teachers and nine English major students were purposively included. In the view of a mixed research design, both quantitative and qualitative data were obtained through document analysis, and teachers and students interview techniques. The document analysis focused on six language domains in the course such as writing, listening, reading, speaking, grammar and vocabulary. The interview focused on the challenges that the teachers confronted and the strategies they used to promote the positive effects of content validity of their assessments in the course. Percentage, Mean Ranking and Pearson Sidney Siegel's Contingency Coefficient were employed to analyze the quantitative data. The qualitative information was also thematically described. The overall result of the study showed that the content validity of the teacher-made assessments was very low (0.211) which in turn reveals negative effects of the assessments on the teaching-learning process of the course. Complex components and demanding nature of the course, student character, constraints and malfunctioned resources as well as teachers' behaviour were the major challenges to properly implement assessment in the course. Therefore, the solution to the problem lies in bringing about improvement on these challenges.

**Keywords:** Challenges; Content validity; Effect; Strategy; Teacher-made assessment

---

Licensed under a Creative Commons. Attribution-NonCommercial 4.0 International License.



\*Corresponding author. E-mail: maldamercy@yahoo.com

ISSN 2521-2192 (Print)

Haramaya University, 2018

## 1. Introduction

The term validity refers to whether or not the assessment measures what it aims to measure (Khan, 2013; Weir, 2005; Wolming and Wikstrom, 2010). It reveals the relationship between language teaching and assessment (Hughes, 2003). Various scholars have attempted to describe the relationships between language teaching and assessment (Thaidan, 2015; Djuric, 2008). A valid assessment represents the teaching materials that are used to teach students in a given subject area (Brown, 2012). The relationship between assessment and teaching is referred to as *systemic validity* (Creswell, 2009), *assessment impact* (Brown, 2012), construct validity, predictive validity, face validity and content validity (Khan, 2013) *consequential validity* (Anderson and Bachman, 2006), *backwash* (Hughes, 2003).

This study mainly focused on the content validity because content validity of assessments directly influences the quality of education (Thaidan, 2015). Moreover, nobody can simply judge the status of content validity of an assessment as good or bad without scientific investigation (Green, 2007). The status of content validity, which depends on the match or mismatches between the contents of assessments and teaching materials, implies either positive or negative effect of the assessment on the language teaching-learning process of a course (Hawkey, 2006). More importantly, content validity investigates the effects of classroom teacher-made assessment in particular on the teaching-learning process of the course (Taylor, 2005). In other words, content validity of an assessment influences the activities of teachers and students, the selection and utilization of teaching materials and the attainment of the objectives of a course (Guma, 2010). For these reasons, Hughes (2003) and Thaidan (2015) argue that teachers should frequently check the content validity of their assessments to enhance the positive effects of their assessments.

Various studies have globally and locally been conducted to demonstrate the importance of content validity of various examinations. Globally, Sukyadi and Mardiani (2011), Shih (2007), Thaidan (2015) and others have studied the content validity of English National Examination at different sites. Locally, Nigussie (2002), Simachew (2012), Yasin (2014) and Hailu (2015) studied the effects of University Entrance Examination, Ethiopian General Secondary Education Certificate in English examinations and the content validity of EFL teacher-made tests at a preparatory school, respectively. The studies exhibited contradicting results. A very few of them found out good content validity (Ching, 2008); whereas, others reported very low content validity of the examinations. The studies concluded that teachers and students' activities are sensitive to content validity of assessments. Therefore, the works recommend that teachers' assessment must cover all the language domains that are embedded in one specific syllabus to attain the objectives of the target course or the unit taught to students (Taylor, 2005).

Specifically, the validity of EFL teachers-made assessment is more debatable than others particularly in Communicative English Skills Course (CESC) (Iyer, 2015). Teacher-made assessment refers to both continuous assessments (CA) or formative assessment (FA) and summative assessment (SA) carried out by teachers as a central issue of teaching-learning process (Anderson and Bachman, 2006). Thus, content

validity of teacher-made assessments in CESC is debatable for three reasons (Taylor, 2005). First, teacher-made assessment is not as valid as the standardized test. Second, the practice and the nature of CA can affect the content validity of teacher-made assessment in CESC. Third, the demanding activities in the domains of the course such as speaking, writing, reading, listening, grammar and vocabulary can construct negative effect of the teacher-made assessment. However, the status of the content validity of teachers-made assessment depends on the degree to which the assessment incorporates a representative sample of the content of whatever objectives or specifications the assessment is originally designed to measure (Taylor, 2005).

On the other hand, content validity of teacher-made assessment in CESC implies several important notions for the measurement of the language objectives that are supposed to be achieved by learners (Taylor, 2005). In this regard, Ching (2008) argues that content validity of teacher-made assessment in CESC is not merely counting the number of contents in an assessment to compare the results against the contents in the teaching materials. Arguably, the degree of representativeness and comprehensiveness of the assessment items with the language domains observed within the teaching-learning materials implies the quality of education (Weir, 2005; Hughes, 2003). In relation to this, Taylor (2005) posed the following eight questions:

...do the participants understand the purpose(s) of the test and the intended use(s) of the results? Are the results provided in a clear, informative and timely fashion? Are the results perceived as believable and fair by the participants? Does the test measure what the program intends to teach? Is the test based on clear goals and objectives? Is the test based on sound theoretical principles which have current credibility in the field? Does the test utilize authentic texts and authentic tasks? Are the participants invested in the assessment activity? (pp. 276-277).

The responses to these eight questions clearly show that content validity is an important issue that all English language teachers failed to do particularly in CESC (Taylor 2005; Brown, 2012; Iyer, 2015). Evidently, Brown (2012) asserts that most (79%) of EFL teachers are unsuccessful to ensure the representativeness of their assessment items to the language domains in CESC teaching materials. Brown (2004) also argues that an assessment with positive effect is generally valid and the reverse is also true.

Green (2007) argues that investigation of content validity is more specific and relevant to teacher-made assessments of CESC because it measures a limited scale of the subject area in a very helpful way to depict the objectives envisioned by teacher (Thaidan, 2015; Khan, 2013). Frequent investigation of the content validity of teacher-made assessments is supposed to determine the effects of the assessments and thereby, enhance the quality of teaching-learning process. Shih (2007) argues that investigation of the content validity of EFL teacher-made assessment is the worthiest instrument in the hands of teacher. This is because content validity mainly affects the feeling and motivation of teacher and students towards the teaching-learning process (Hughes, 2003; Iyer, 2015). Consequently, whatsoever the effect of the assessment is, practitioners should have a clear understanding on how their assessments affect the quality of education in the course (Iyer, 2015; Taylor, 2005).

The researcher's personal experience shows that the issue of content validity of teacher-made assessments in CESC involves more serious problem in Ethiopian in general and in Ambo University in particular. The assessment trend in the department of English Language and literature of Ambo University in particular was changed during 2016 to 2018. Unlike the trends before 2016, CA, which accounts for 60% of students' grade in CESC, is usually carried out by teachers individually, but SA, which accounts for 40% of their grade, has been prepared by course coordinating committee established at department level. In addition to the assessment trends, three different teaching modules were used from 2016 to 2018. Therefore, the ultimate goal of this study was to internalize the problem because a successful assessment executer is more likely to be able to meet the increased teaching demands as a University teacher (Davies, 2014). Hence, building on the content validity framework, this study investigated the content validity of EFL teacher-made assessments, the challenges EFL teachers faced and the strategies they used to promote the positive effects of their assessment in CESC during 2016, 2017 and 2018 at Ambo University.

This study evaluated the content validity of teacher-made assessments of CESC during 2016 to 2018 at Ambo University. This is because the validity of every teacher-made assessment of CESC is questionable everywhere in the world (Taylor, 2005; Brown, 2012). Supporting this idea, Taylor (2005) argues that content validity of teacher-made assessment in the course is less valid; therefore, it should frequently be checked (Hughes, 2003). Iyer (2015) also argues that content validity is an essential concern that all English language teachers failed to achieve particularly in CESC. Brown (2012) asserts that most (79%) of English language teachers are ineffective in ensuring the validity of their assessment contents in CESC. This implies that the problem of the content validity of EFL teachers-made assessments should be of concern to everyone who teaches the course. In relation to this, Davies (2014) argues that validity of teacher-made assessments is the base for the quality of teaching-learning process because it affects the classroom activities of both teachers and students.

The issue of content validity of EFL teacher-made assessments in CESC involves more severe problem in Ambo University in general and in the department of English Language and Literature in particular. The researcher's experience as an EFL teacher in the University reveals that the change in assessment culture has created uncomfortable situation in teaching learning process of CESC during 2016 to 2018. As CA was carried out by individual teachers, the practices of CA were frequently observed to be varying from person to person in assessing the same objectives of the course. These inconsistent assessment practices had created grievances among the teachers and the students on the fairness of the students' grade. This grievance had in turn frustrated the teachers to determine reasonable grades for their students. More seriously, 25 (28.4%) of 88 students have scored less than 'C' grade within 2016 and 2017. The assumption behind this scenario indicates that both forms of teacher-made assessments fail to measure the entire components of CESC (Ambachew, 2003; Motuma, 2017).

On the other hand, the issue of the content validity of teacher-made assessment has rarely and partially been studied in Ethiopian context. For example, Hailu (2015) has tried to determine the content validity of EFL teacher-made tests of grade 11 at Ambo preparatory school. His study was incomprehensive because he did not include the content validity of all language domains and the analysis of CA in the school. Moreover, although Ethiopian Universities use the harmonized CESC curriculum and have similar directives, policies, rules or regulations of assessment procedures, the problem of content validity of EFL teacher-made assessments has not been determined. Specifically, knowing this severe and pressing concern, no investigation has been conducted to evaluate EFL teacher-made assessment content validity in CESC at Ambo University. This implies that EFL teacher-made assessment validity has been given less emphasis particularly in such a complex course that requires demanding efforts to manage. Consequently, this study has intended to address the following research questions:

1. To what extent do the three years (2016, 2017, and 2018) communicative English skills course teacher-made assessments reorient the teaching materials?
2. What are the challenges that EFL teachers confront to promote the validity of their assessment contents in communicative English skills course during the respective years?
3. What are the strategies used by EFL teachers to enhance the validity of their assessment contents in the domains/components of CESC?

## **2. Research Methodology**

### **2.1. Research Approach and Design**

This study was grounded on pragmatism as a research paradigm. The rationale behind choosing pragmatist paradigm is the nature of the topic, the problem and the research questions of this study (Creswell & Plano Clark, 2011; Davies, 2012b). In alignment with the pragmatist view of knowledge, a mixed research design was employed (Davies, 2012a, 2014). The benefits of using a mixed methods approach are to address the research problem and research questions (Creswell & Plano Clark, 2011). The research questions of this study require the use of both quantitative and qualitative approaches because the use of either qualitative or quantitative approach cannot comprehensively address the problem (Creswell, 2014; Motuma, 2014). The use of both approaches ensures the opportunity to include objective and subjective data (Creswell, 2009; Creswell & Plano Clark, 2011). Besides, mixed research approach help to provide both quantitative and qualitative answers to the questions of “who”, “what”, “why”, and “how” (Creswell & Plano Clark, 2011).

In view of mixed research design, a descriptive correlational research method was employed to evaluate the relationship between EFL teacher-made assessments and the teaching-learning processes of CESC. This method was used to answer the questions as well as to assist the researcher to determine the sample and sampling techniques, data gathering tools and data analysis methods.

## **2.2. Data Sources, Sample Size and Sampling Techniques**

The major sources of the data were CESC teaching materials (modules), SAs and CAs made by EFL teachers from 2016-2018 at Ambo University. Three years SAs and CAs and two major modules were analyzed to determine the degree of the content validity of EFL teacher-made assessments. Principally, two major modules were used to teach CESC from 2016 to 2018. The first harmonized module was used from 2011 to 2016 with minor modifications that have been made every year by the course coordinating committee. The second module was used only in 2017 E.C. In 2018, the previous module was used again with minor amendments from its parts.

In addition to the documents, EFL teachers and students purposively participated in the study to include judgements of expertise and opinions of the beneficiaries (Cresswell, 2014). According to Cresswell (2014), expert judgement is the most important method researchers use to determine whether an assessment has good content validity or not (Cresswell, 2014). For this reason, 12 of 32 full-timer Ambo University EFL teachers were purposively selected. This is because three of them have been teaching CESC to English major students during the time of study, and nine of them worked as the members of course coordinating committee. The assessment committee has been revised every year, but some of the committee members have worked for more than one year. For similar purpose, nine students (three from high, three from medium and three from low scorers in their CAs results) purposively participated (Creswell, 2009). This is because Creswell (2014) symbolically explains that students should be included in such a study because they are the fertile land on which a farmer sows the seeds and harvest the products later.

## **2.3. Data Collection Methods**

To measure content validity of the assessments, document analysis, teachers' semi-structured interview and students' pre and post SA interview (Ching, 2008) were used to collect data for the study. With regard to document analysis, Shannon and Hsieh (2005) point out that content analysis currently incorporate three distinct approaches: conventional, directed, and summative. From the three approaches summative content analysis was employed in this study. According to Shannon and Hsieh (2005: 1283), "Summative content analysis is fundamentally different from the prior two approaches because it helps to analyze data in relation to particular content and context". Moreover, summative content analysis involves both quantitative and qualitative data analysis to determine the alignment between the assessments and modules of CESC (Cresswell, 2014; Okeeffe, 2013). To this end, a checklist and a protocol analysis were developed based on the suggestions of Okeeffe (2013). Moreover, three EFL teachers participated in the development of the instruments to check the validity of the document analysis.

In addition to document analysis, semi-structured interview was developed and conducted on 12 EFL teachers based on the recommendations of Kothari (2004). The purpose of the interview was to gather opinions (expert judgements) of the teachers on the status of their assessments' content validity, the challenges the teachers faced and the strategies used by EFL teachers to promote positive effects of their

assessments on teaching-learning process of CESC (Cresswell, 2014; Cohen, Manion & Morrison, 2007). The interview was piloted with three EFL teachers who were not included in the actual study as the source of data to check the validity and reliability of the instrument. Moreover, pre and post SA interview was also conducted with nine students to supplement the data obtained through document analysis and teachers' interview (Kothari, 2004).

#### 2.4. Data Analysis

Both quantitative and qualitative data analysis methods were employed to determine the content validity of the assessments based on Cresswell's (2014) recommendations, to describe the challenges EFL teachers face and the strategies they use to maintain positive effects of their assessments on teaching-learning of CESC. The data obtained through document analysis were quantitatively analyzed using descriptive statistics (percentage, mean ranking) and Pearson Sidney Siegel's correlation contingency coefficient 'C' to make judgments on the content validity or effects of the assessments. The content analysis followed three sequential steps: analyzing the contents in teaching materials, analyzing the contents in assessments and determining the level of relationship between the assessments and the teaching materials using Siegel's contingency coefficient (C). The contents of the modules and the assessments were similarly organized under six major components: Speaking, Reading, Vocabulary, Grammar, Listening, and Writing. To calculate the Siegel's contingency coefficient 'C', the following formula and interpretations were used.

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

Where:

C= the value of relationship

N=grad total

X<sup>2</sup>= refers to Chi-square

To determine value of Chi-square (X<sup>2</sup>) from the summary of contingency table, the formula is:

$$x^2 = \frac{(\textit{observedvalue (o)} - \textit{(E)expectedvalue})^2}{\textit{expectedvalue}}$$

The values of Sidney Siegel's contingency coefficient are interpreted (Green, 2007) as indicated in Table 1.

Table 1. Interpretation of the effects of content validity

No.	Value of contingency	Descriptions and interpretations	Content validity	Effects
1	C= 1	Perfect correlation	Valid	Positive
2	C= 0.91-0.99	High correlation, proportional	Valid	Positive
3	C= 0.71-0.91	Moderate correlation	Acceptable	Negative
4	C= 0.41-0.70	Low correlation	Unsound	Negative
5	C= 0.21-0.40	Very low correlation	Worthless	Negative
6	C= > 0.2	No correlation, irrelevant,	Unacceptable	Negative

Source: Green (2007)

The data obtained through teachers and students' interview were qualitatively analyzed. The analysis of the qualitative data was made based on the theme-based analysis (Cresswell, 2014). Hence, the challenges that the teacher faced to improve the content validity of their assessments were sorted into the complex and demanding nature of CESC, teacher activities, student character and resource related challenges. Similarly, the strategies used by EFL teachers to enhance the positive effect of their assessments of CESC were also categorized into test design strategies, teaching-learning strategies and classroom resource management strategies.

### 3. Results and Discussion

#### 3.1. Results

This section incorporates three sub-topics. The first is the analysis of the course materials and the assessments analysis to determine the level of relationship between the assessments and the course materials using descriptive statistics and Siegel's contingency coefficients. The second section is about the analysis of challenges EFL teachers faced and the strategies they used to make their assessment contents valid. The final sub-topic presents the discussion of the results. Hence, the following section presents the data analyzed on the contents of the course materials from 2016 to 2018.

##### 3.1.1. The result of course materials analysis

Course materials analysis was the main source of data to determine the alignment of the course with the assessments. Thus, two major modules and one revised version of the previous module were analyzed. This is because the two major modules were used during 2016 and 2017, and the amended version of the previous module was employed to teach the course. All modules have five units. Every unit of the modules in turn has six major language domains. The sequence of the components of the modules varies from chapter to chapter within a module. Table 2 presents summary of contents in every language domain in the teaching materials.



Table 2. Frequency of the contents of the course materials from 2016 to 2018

No.	Content area	Frequency in each year			Total frequency	%	Rank order
		2016	2017	2018			
1	Speaking	27	53	27	107	10%	5
2	Reading	119	121	119	359	33.4%	1
3	Vocabulary	57	86	57	200	18.6%	2
4	Grammar	48	86	48	182	17%	3
5	Listening	59	62	59	180	16.8%	4
6	Writing	18	9	18	45	4.2%	6
	Total	328	417	328	1073	100%	

Source: Communicative English Skills course modules

Table 2 shows the frequency of contents in each component and the cumulative frequency in CESC modules by percentage and by rank order. The number of items in the modules is 1073 during 2016 to 2018. The maximum number (417) of frequency of contents in CESC was observed in 2017. The distribution of language contents in the modules is reading (33.4%), vocabulary (18.6%), grammar (17%), listening (16.8%), speaking (10%) and writing (4.2%) skills items in descending order. This demonstrates that reading was the dominant, but, writing was the least language component in the modules throughout the years. The following section presents the items in CAs and SAs in CESC during 2016 to 2018.

### 3.1.2. Analyses of the assessments

This section discusses teacher-made assessments in each component of CESC by year. The assessments have two categories: continuous assessment (CA) or formative assessment (FA) and summative assessment (SA) items. In order to determine the level of the representativeness of the assessments, the items in CA and SA were sorted into six major components of the language domains in CESC based on the language domains in the modules as summarized in Table 3.

Table 3. Frequency of the contents of the items in the assessments from 2016 to 2018

Language domains	Frequency in each year						Total frequency				Rank
	2016		2017		2018		CA	SA	T	%	
	CA	SA	CA	SA	CA	SA					
Speaking	-	6	1	8	1	6	2	20	22	9.6	4
Reading	11	15	10	12	12	20	33	47	80	34.7	1
Vocabulary	3	4	7	14	6	10	16	28	44	19.12	3
Grammar	12	15	11	16	13	12	36	43	79	34.4	1
Listening	-	-	-	-	1	-	1	-	1	0.44	6
Writing	2	-	1	-	1	-	4	-	4	1.74	5
Total	30	40	30	50	32	48	92	138	230	100%	

Source: Assessment booklets of the years 2016-2018

Table 3 presents a comparison of assessment items among the language components by year. The table indicates that a total of 230 assessment items of CESC were set within the three years. Specifically, a total of 130 SA and 92 CA items were carried out by the teachers during the years. More specifically, the maximum number of SA items (50) was set in 2017, but 48 and 40 SA items were set during 2018 and 2016, respectively. The number of CA items seems to be uniform throughout the years. However, the distribution items between and within the assessments were disproportionate; for instance, the distribution of vocabulary items were intermittent from year to year: 7 in 2016, 21 and 17 items in 2017 and 2018 respectively. Moreover, reading and grammar (69.1%) dominated the assessments, but listening and writing skills were overlooked by the teachers throughout the years. Similarly, the result of interviews asserted that the students were not adequately assessed in listening and writing skills even in CAs.

### 3.1.3. Relationships between course module and assessment

This section demonstrates the analyses of the documents to determine the relationship between the assessments and the modules through the comparison of three determinants: the contents in the modules, the time allotted to cover the contents and the items in the assessments. Although some language domains in the modules were either too general or overlapping with each other to exactly determine the time allotted to cover each language domains, efforts were made to determine the time allotted to the domains. To make the time allotment as closer as possible to the contents, three EFL teachers' expert judgements were used in addition to the researcher. The analysis of the proportion of the time allotted in relation to the six major language content areas is indicated in Table 4.

Table 4. The relationship of modules and assessments in relation to time allotment

No.	Language domains	Total frequency of course module		Time allotted to language domain		Total number of items in assessment	
		F	%	F	%	F	%
1	Speaking	107	10%	7 hrs	14.58	22	9.6
2	Reading	359	33.4%	9 hrs	18.75	80	34.7
3	Vocabulary	200	18.6%	8 hrs	16.67	44	19.12
4	Grammar	182	17%	8 hrs	16.67	79	34.4
5	Listening	180	16.8%	7 hrs	14.58	1	0.44
6	Writing	45	4.2%	9 hrs	18.75	4	1.74
		1073	100%	48 hrs	100	230	100

Source: Module (2011 & 2017) and Assessment materials (2016, 2017, 2018)

Table 4 reveals the amount of time allotted to each language domain in the modules by year. The table indicates that speaking, reading, vocabulary, grammar, listening, and writing were given 7 (14.58%), 9 (18.75%), 8 (16.67%), 8 (16.67%), 7 (14.58%) and 9 (18.75%) hours, respectively, from the total of 48 hours as CESC. The amount

of time given to each domain seems to be almost proportionate to the domains. However, referring to both CA and SA, 159 (69.1%) questions were set from reading and grammar sections only; whereas only 25 (11.78%) items were included to the assessment from listening, writing and speaking skills. Comparing with the time estimated to cover the reading objectives, unreasonable reading items were prepared. This shows incompatibility between the time estimated to the skills and the assessment components, as illustrated in Figure 1.

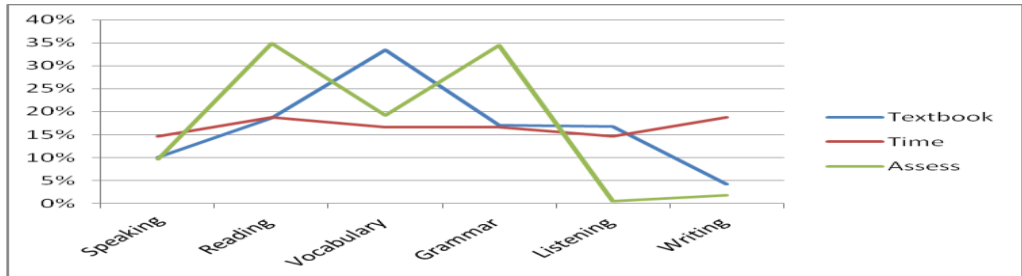


Figure 1: Relationship between module and assessment in relation to time allotted  
 Source: Module (2011 & 2017) and Assessment materials (2016, 2017, 2018)

Figure 1 above shows estimation of time allotted to cover each of the six major language domains in the modules. The figure signifies that, totally, 33.3% of the 48 hours was allotted to writing and listening sections, whereas only 2.18% of the questions was included into the assessments from listening and writing skills. This reveals that the contents in the assessment items were not comprehensive and proportionate to the attention given to them in the modules. For this reason, both teachers and students assumed that the time allotted to listening and writing was wastage. Table 5 presents the summary of contingency. In the table, the abbreviations such as: *TbOV* means textbooks observed value; *EOV* means examinations observed value; *TOV* means total observed value; *TEV* means textbooks expected value; *EEV* means examination expected value.

Table 5. Summary of contingency: content in the assessments and the modules

No.	Contents	TbOV	EOV for each year			TbEOV	EEV for each year		
			2016	2017	2018		2016	2017	2018
1	Speaking	107	8	10	7	132	4.8	5.83	4.83
2	Reading	200	26	21	31	278	7.6	9.49	8.88
3	Vocabulary	359	7	21	16	403	14.2	14.05	15.27
4	Grammar	182	27	27	24	260	7.2	10.22	7.19
5	Writing	180	2	1	1	184	5.1	4.138	6.93
6	Listening	45	-	-	1	46	0.98	0.68	1.797
	Total	1073	70	80	80	1303			

Source: Computed from the previous tables and figures

The figures in Table 5 show the sum of the total expected value (1303) of the teaching materials, which were computed from the total observable value (1073) and observable values (230) in the assessments. In other words, TbEOV is the summation of EOV and TbOV, which, means for example, the grand total of 1073 (EOV) and 230 (EOV) is 1303 (TbEOV). To determine the value of Chi-square ( $X^2$ ) from the summary of contingency table, the formula is:

$$x^2 = \frac{(\text{observedvalue } (o) - (E)\text{expectedvalue})^2}{\text{expectedvalue}}$$

Finally, based on the value of Chi-square ( $X^2$ ), Sidney Siegel’s correlation coefficients were determined. Table 6 summarised the results.

Table 6. Summary of Sidney Siegel’s correlation coefficients and their interpretation

S/N	Years	Relationship by years			Interpretation of the content validity and its effects of assessments by year
		N	$X^2$	C	
1	2016	398	64.149	0.42	Low correlation; worthless effect
2	2017	497	13.1133	0.385	very low correlation; negative effect
3	2018	408	31.0012	0.573	Low correlation; worthless effect
4	Total	1303	65.516	0.211	Very low correlation; negative effect

Source: Computed from the previous tables and figures

The data in Table 6 reveals the summary of Sidney Siegel’s Correlation Coefficients (C) by year and their interpretation. The table indicates that the degree of relationship between the contents of the modules and the assessments found to be  $C = 0.42$ ,  $C = 0.385$  and  $C = 0.573$  in 2016, 2017, 2018 respectively. The overall result of Sidney Siegel’s contingency coefficient ( $C = 0.211$ ) reveals that the correlation between assessments made by EFL teacher during 2016-2018 and the modules is very low. This disproportion between the contents of the modules and assessments imply harmful effects of the assessments in CESC. The result of the students’ interview also reveals similar phenomena in which they complained their teachers’ assessment practices that they assess what their students had not learnt. This in turn affected motivation and feeling of students towards the utilization of the modules.

**3.1.4. Challenges related to teacher-made assessments**

This section presents the analysis of challenges the teachers faced to promote the positive effects of their assessments made during 2016-2018. To this end, efforts were made to categorize the challenges under four major themes: the demanding and complex nature of the course, teachers’ activities, learners’ character and resource constraints.

To begin with, the demanding and complex nature of the course was identified as the most critical challenges by the teachers. The teachers asserted that the multiple domains of language objectives in CESC created a mismatch between their assessments and the modules. This is because teachers faced difficulties to include all the language domains in their assessment from CESC components. Similarly, the

complexity of designing, constructing and administering all assessment strategies, correcting and measuring students' results were the most common demanding activities for the teachers to make their assessments comprehensive and proportionate. In relation to this, it is an evidence to quote the voice of a teacher that might represent others' opinion on their perception about CESC.

Communicative English Skills course involves multifaceted objectives and contents. It is difficult for me to assess these all language domains at a time. Moreover, when I construct the assessment items, how should I worry about the proportionality and comprehensiveness of the language domains in the course and the time allotted to each? Moreover, it is difficult for me to align the language domain with its appropriate assessment techniques. Even if I try to implement the assessment strategies, how can I correct all items set from different language domains and feedback for the students?

In this quotation, one can infer that complexity and the demanding nature of CESC are the critical challenges the teachers face to align their assessment techniques within its language domains. The teachers were also challenged by the multifaceted objectives and the demanding activities in correcting and giving feedback to the students to make their assessment comprehensive and proportionate to the objective to be assessed in CESC. The analysis made from the teachers and the students' interviews so far support the results of the analysis of the documents. The analysis shows that the teachers were less strategic to overcome the complexity and the demanding nature of the assessments. As a result, both the students confirmed that they did not carefully study the modules; they usually review the previous examinations items because they knew that their teachers repeat the preceding items. Almost all (seven) of the nine students who participated in the interview also complained that their teachers "compile the items from elsewhere".

Resource constraints were the second serious challenges for teachers to include all the components of language domains into their assessment. The teachers explained that as CESC is a multidimensional, it requires various resources to assess its components in a comprehensive way. However, constraints of reference materials, shortage of computer, lack of cameras, very small size and poorly furnished laboratory, frequent internet interruption, problem of duplicating machines and stationery materials have challenged the teachers to enhance positive effects of their assessments. Moreover, large class size and constraints of time created due to instability Ambo University have affected the teachers' assessment activities. The students also criticized that they rarely went to language laboratory to learn, but they were not assessed using any sort of technological devices.

Student character was the third most emphasized challenges by the teachers. Student character as a challenge of content validity of the assessments in CESC encloses the preoccupied learning and assessment culture, ethnic diversity, disciplinary problems, language deficiency, cheating in collaborative learning, lack of intrinsic motivation and negative attitude toward the use of CA, self and peer assessment in CESC. The teachers also evaluated that these affected students' self-initiative and self-reliance, self-evaluation, goal setting practices and appropriate reaction to their teachers' assessment techniques in CESC. The students confirmed

similar connotations with the teachers' expert judgments that they were reluctant to be assessed through self and peer-assessments.

Teacher character was solemn challenge for the content validity of their assessments. The questions were asked to respond to how they designed assessments and how they managed student character of as well as resources constraint they need to carry out assessments. They illustrated that they design assessments without guideline and table of assessment specification. However, they internalized that teachers' prior experience, poor pedagogical skills and low interest towards CA challenged their assessment activities. In relation to teacher character as a challenge, a teacher said the following.

I do not believe that I should include all the language items in one assessment. Two or three language domains are enough for an assessment. Students will learn writing skills by the next semester in their basic writing skills. Moreover, I skip over listening activities because there is no efficient laboratory. I know that summative assessments have been designed by course coordinating committee, but most of the newly graduated teachers are not effective in designing comprehensive and proportionate continuous assessment items in particular. Others also fail to reverse the traditional assessment paradigm they have experienced in other courses.

This summarizes that the teachers failed to design valid, clear, practical, reliable and measureable assessment techniques in CESC. Teacher character and their perception challenged them to promote validity of their assessment contents. The teachers skipped over listening and writing during teaching and assessment in CESC. This is because they knew that students could have chance to separately exercise writing skills more in the upcoming basic writing skills course. This implies that they give less emphasis to some language skills but more weight to others. In general, it is possible to deduce that teachers' inability to handle students' interest and the complex nature of CESC challenged the content validity of teachers' assessments.

### **3.1.5. Strategies used by EFL teachers to promote positive assessment effects**

This portion highlights the analysis of the plausible and recommended strategies used by EFL teachers to enhance their assessments' content validity. EFL teachers' strategies to enhance positive effect of their assessments in CESC were categorized into assessment design, teaching-learning and classroom resource management strategies. These strategies are normally supposed to align assessments to modules and thereby, enhance positive effects of the assessments.

In relation to assessment design strategies, the responses of EFL teachers in their semi-structured interview revealed that the teachers use the module only to design their assessments. Even the assessment committee did not have any guideline and a pre-determined table of assessment specification. This led them to implement CA differently in particular. Moreover, there were variations among teachers in the implementation of SA because some teachers do not complete the course. These discrepancies imply negative effects of assessments on the teaching-learning of CESC.

The alignment of assessment strategies to teaching-learning strategies can enhance positive effects of teachers-made assessments. To ensure this alignment, most teachers think to assess what they have taught. However, they asserted that

sometimes set assessment items from elsewhere because they normally believe that language assessment should not totally be confined to the course material. On the other side, some other teachers argued that setting assessment items from elsewhere result in students' unfair grade in CESC. Besides, they complained that there was mismatch between the teaching-learning and assessment strategies because teachers teach CESC in communicative language approach, but they assess it using traditional testing approach.

Proper resource utilization and management strategies necessarily promote positive effects of EFL teacher-made assessments. However, EFL teachers were ineffective in resource utilization and management. They failed to manage time constraints created due to instability in the University. Most of the teachers were also reluctant to use alternative resources, for example, teachers' loud reading to assess listening skills and chalkboard to assess other language components instead of waiting for scarce resources to carry out the assessments. Both teachers and students considered the teaching-learning of listening skills without listening devices is boring and challenging.

#### **4. Discussion**

The discussion section integrated the findings identified through document analysis, teachers and students' interviews. It discusses the status of content validity, the challenges EFL teachers faced and the strategies they used to promote positive effects of EFL teacher-made assessments in CESC from 2016-2018 at Ambo University. The review and the analysis of the course module indicate a list of general objectives and specific objectives. The review of course materials mainly imply that the objectives CESC is fundamentally different from other language course because the course integrates complex language domains to be assessed using different type of items, tasks and techniques (Burger, 2008; Davies, 2008).

In relation to the analysis of documents, the objectives of the course as stated in the teaching materials were designed to develop students' speaking, reading, listening, and writing skills including grammar and vocabulary. Compared to the teaching materials of the course, reading, grammar and vocabulary items predominantly appeared in the assessments in descending order. In other words, the assessment items did not proportionally incorporate some language skills of the course such as writing and listening skills items, even in CAs, for various challenges including the malfunctioned language laboratory and scarce resources.

The degree of the relationship between the contents of the teaching materials and the assessments found to be very low (0.211) or only 21.1% valid. This shows negative effects of the assessments on teaching-learning process in CESC. Specifically, the Correlation Coefficient values "C" is 0.42 for 2016, 0.385 for 2017 and 0.573 for 2018 which were evidences for EFL teacher-made assessments' to have negative effects. However, the values of the correlation coefficient indicate insignificant progress of positive content validity of EFL teacher-made assessment from 2016 to 2018. Evidently, no speaking and listening skills items were included

into CA in 2016, but only 2 (1%) of CA items were included from listening and speaking skills in 2018.

The mismatch between the contents of the course materials and EFL teacher-made assessments was apparent throughout the years. For example, in the course materials, reading, vocabulary, grammar, listening, speaking and writing constituted 33.4%, 18.6%, 17%, 16.8%, 10% and 4.2%, respectively. However, reading, vocabulary, grammar, listening, speaking and writing covered 34.7%, 19.12%, 34.4, 0.44%, 9.6% and 1.74%, respectively, in the assessments. These figures show that except the vocabulary, speaking and reading skills, other language items in the assessment do not reorient the distributions of contents in the modules. Evidently, the intersected lines in Figure 1, clearly illustrate the distribution of the contents in the modules, the time allotted to cover the contents in the module and contents in the assessments.

Comparing with the attention given to each language domain, the time allotted to cover each of the language domains in CESC seems to be reasonable. Evidently, 7 (14.58%), 9 (18.75%), 8 (16.67%), 8 (16.67%), 7 (14.58%) and 9 (18.75%) hours were estimated to cover the speaking, reading, vocabulary, grammar, listening, and writing portion respectively. However, the distributions of items in assessments were not proportionate to the attention given to the objective in CESC module. Such disproportionate distribution of contents was an indicator of negative effect on teaching-learning process of the course during the years (Hughes, 2003; Cheng, Watanabe & Curtis, 2004).

Various scholars including Hughes (2003) and Cheng *et al.* (2004) suggest that all the major language domains from the course material are expected to be assessed either by CAs or by SAs. However, the reality in CESC during the sample time at Ambo University deviated from what these scholars have suggested. For instance, in the assessments, SAs included only one item from listening and four items from writing throughout the years. The highest portion 47 (34.1%) of SAs and 33 (35.9%) of CAs was occupied by reading. In similar context, 43 (31.2%) of SA and 36 (39.1%) CA items were taken by grammar. These evidences, therefore, confirm the argument that the distributions of EFL teachers-made assessment items in CESC were incomprehensive and disproportionate to the teaching materials (Iyer, 2015).

Concerning CA, although the overall intention of CA is to involve students in a variety of activities as stipulated in the objectives of the course materials to promote the students' language skills, the content validity of EFL teacher-made CAs did not have significant difference from the SAs. For example, only 5 (2.18%), one from listening skill and four from writing skill, items were included to CA items throughout the years, but 69 (75%) of CA items were from reading and grammar. This indicates that CAs did not represent the contents in the module. The time allotted to listening and writing was considered as wastage by the teachers and the students because students' listening and writing skills were not adequately assessed even in CA.

The teachers were required to reply to a question why the content validity of their assessments in the course is low. They listed several challenges that had confronted them to promote positive content validity of their assessments. The challenges were



classified into four major themes based on the suggestions of Iyer (2015) and Jabbarifar (2009). The major themes of the challenges comprise the demanding and complex nature of the course, teachers' behavior, learners' character and resource constraints.

With regard to the complexity and demanding nature of the course, EFL teachers explained that CESC involves demanding activities to assess the language domains in a comprehensive way because the course encloses complex language components. These demanding activities and complex language domains in the course were identified as the critical challenges for the teachers. The most common demanding activities of CESC consist of the difficulty of designing, constructing and administering the assessments employing all the assessment strategies, correcting and measuring students' results. Moreover, conducting effective CA was challenging for the teachers because of students' cheating. In other words, they faced difficulties to promote the practicality, validity and measurability of their CA assessment activities and techniques; for instance, home works and group works were challenging the content validity of EFL teachers' assessments for uncontrollable students' cheating. The teachers also admitted that their assessments involved mismatches between the assessments and the teaching materials. Moreover, they justified that they could not prepare inclusive assessment due to the multiple domains of language objectives in the course.

Communicative English skills course, as a multidimensional course, requires a variety of resources to carry out comprehensive and proportionate teacher-made assessments (Five and Nicole, 2013). To state the argument clearly and precisely, the constraints of resources fundamentally challenged the teachers to carry out comprehensive assessments in such a complex course. For example, scarcity of resources was identified as critical challenges for the teachers' activities to promote positive effects of their assessments. Moreover, large class size and time constraints, created due to students' unrest in the University, challenged the teachers' assessment activities.

In relation to the challenges created by resource constraints, the teachers asserted that listening skill in particular has mostly been neglected in both CAs and SAs in CESC because of the problems of listening facilities in language laboratory. However, although it is well expected that the teachers might not include the speaking and listening questions in SAs because of the constraints of laboratory devices and other resources, there is no reason to exclude speaking, writing and listening items from CAs. This is because one can argue that writing skills can directly be assessed using SA. Similarly, speaking skills can also be assessed using SA in the form of dialogues which are common even in standardized tests. Moreover, in the area of CA, teachers are expected to prepare speaking and listening questions in CA and administer it either in classroom or in language laboratory (Hughes, 2003). However, it is possible to deduce that the assessments influenced what the teachers teach, the degree and depth of their teaching, their attitudes towards the content and method of teaching and learning process of the course (Burger, 2008; Motuma, 2015).

Another most concerned challenge by EFL teachers was student character. Student character as a challenge of content validity of the teacher-made assessments in the course encloses the preoccupied students' traditional learning and conventional assessment culture, students' learning styles and ethnic diversity, students' disciplinary problems, their language deficiency, students' cheating in general and in collaborative assessment activities in particular, students perception towards CA, lack of intrinsic motivation and negative attitude toward the use of self and peer assessment in CESC. The teachers also expressed their observations that the students' poor self-initiative and self-reliance, lack of self-evaluation and goal setting practices challenged them to properly implement CA techniques and thereby effectively assess their students in CESC.

On top of others, students' lesser consideration for CA challenged the teachers to properly implement the assessment techniques. The teachers also complained that the students considered CA as it is usually given to support their grades, not to assess their knowledge, attitudes and skills. For instance, the teachers confirmed that students did not worry about the result of their CAs; they usually prepared themselves for SA. To this end, they often reviewed previous examination items to score good result in SA. Moreover, the students gave lesser attention to listening, writing and speaking skills than others even during the time of CA.

Students' language deficiency was also a critical challenge to implement CA in the course. Although CESC requires different alternative assessment techniques, the teachers asserted that they could not use peer-teaching, classroom works, home works and other collaborative activities to assess the students due to the students' language deficiency and cheating. This is because the teachers complained that students are unable to use English language during the listening and speaking classes; rather, they use their mother tongue. As a result, they do not want to work with the students who cannot use their respective language. The teachers were also challenged by the students' cheating to give activities as homework or group works to assess their students' ability because the students do not do it for themselves; unexpectedly, they usually give it to another better student they think among themselves to do it for them. Hence, the teachers think that considering 60% of the students' grade from CA results is giving free grade to many students because they do not perform it. For this reason, implementing proper CA itself in the course is the most critical challenge for the teachers.

Content validity of teacher-made assessments in CESC was also challenged by teacher characters. The teachers failed to design comprehensive assessment and administer it from the course. They were also ineffective to manage the interest and character of students. Besides, they were unsuccessful in managing scarce resources they need to carry out the assessment. Moreover, inefficient teachers' activities in selecting and designing additional relevant, authentic teaching and assessment materials have challenged the content validity of their assessment in the course. Furthermore, teachers' previous working culture, poor pedagogical skills, poor classroom English, low commitment and interest as well as teachers' improper perception towards the role of CA in CESC challenged the content validity of their

assessments. Moreover, teachers' failure to manage the complex components and the demanding activities of CESC were identified as a challenge of the teacher-made assessments.

In order to enhance positive effects of the teacher-made assessments in the course, EFL teachers have been employing a variety of strategies (Five and Nicole, 2013). For example, they mainly used harmonized teaching modules to teach the course to the students to make their assessment practices uniform. Moreover, they used course coordinating committee to design SAs every year in order to avoid unnecessary variations among the teachers in giving grades to their respective students. They also avoided the use of additional materials in teaching the course for similar purpose. Besides, they usually carried out CA individually to avoid complex process of the committee. However, the content validity of EFL teacher-made assessments remained very low throughout the years. As a result, one should ask fundamental questions here. Why does the content validity of EFL teacher-made assessments in the course remain low? How do EFL teachers manage the existing situations in the course so that they can overcome the challenges and ensure the practicality of their assessment techniques? Were the materials they use to assess the students ability authentic and relevant to the choice of the students and to the nature of course? These all questions require further investigations.

## 5. Conclusions

EFL teachers in the university have been employing a variety of strategies to enhance the content validity of the teacher-made assessments. They use individual teacher to design and administer the CAs to simplify the complex process of committee works. However, they used the harmonized course modules to teach the course and coordinating committee to design and process SAs to maintain uniformity among the teachers. They also used collaborative, self and peer-assessment strategies to assess the students' performances. However, the use of these strategies could not enhance the content validity of the teacher-made assessments in the course.

The content validity ( $C = 2.11$ ) or 21.1% of EFL teacher-made assessments in the course was very low throughout the years. This low content validity was resulted from the incomprehensive and disproportionate distributions of the language domains within and among the teacher-made assessments. A mismatch between how to teach and how to assess has also implied negative effect of the teacher-made assessments on the teaching-learning of the course. The modules of the course was designed based on the principles of communicative English skills course, but the teacher-made assessments were dominated by traditional grammar and reading comprehension throughout the three years. This mismatch was created by a variety of challenges including the demanding and complexity of the course, ineffective teachers' performance, improper perception of teachers and students towards CA, negative attitude and low motivation of teachers and students towards CA, uncreative character of learners and, constraints as well as poor resource management.

## 6. Recommendations

Embarking on the harmful effect of their examinations, EFL teachers should first analyze the existing course and student related situations carefully to ensure the practicality and the validity of their assessment techniques.

The teachers should develop a guideline to design uniform CAs in very clear and precise terms. Based on the guideline, the teachers should determine the authenticity and relevancy of the materials they use to assess the students' ability to fit the materials to the interest and motivation of the students and the complex nature of the course.

They were also recommended to carefully draw up table of specification to make their items comprehensive and proportionate within and among their assessments.

EFL teachers should employ relevant and practical assessment design strategies, relate the assessment strategies to the teaching-learning strategies and thereby, properly employ the classroom resource management strategies to promote the content validity of their assessments in the course.

To this end, teachers are expected to up-date and equip themselves with these skills through self-training and conducting action research or problem-solving studies.

Department of English Language and Literature, College of Social Science and Humanities and Ambo University should work together to give trainings to EFL teachers to alter the teachers' perception, attitude and motivation and thereby, implement a various assessment techniques.

Department of English Language and Literature should revise its assessment policy to reduce the weight of CA but to increase the weight of SA to maximize the credibility of students.

Ambo University should also minimize the class size for English language teaching, furnish the language laboratory and provide materials so that EFL teachers can assess all the necessary language domains they want to assess including writing and listening skills.

Researchers in other sites can conduct studies on the effectiveness of the continuous assessment in detail.

## 7. Acknowledgements

I render my thanks to God for His absolute love and protection. I appreciate my wife Mikire Dassie and my children: Malda, Mercey, Mirre, Miracle and Malgarin Motuma for their encouragement.

## 8. References

- Ambachew, M. D. 2003. *A review of the Ethiopian related literature on reading ability*. University of Pretoria Ltd. Pp. 83-93.
- Anderson, C. J. & Bachman, F. L. (Eds.). 2006. *Assessing young language learners*. Cambridge, New York: Cambridge University Press.
- Brown, D. 2004. *Language assessment*. San Francisco State University: Longman.
- Brown, J. D. 2012. Choosing the right type of assessment, In C. Coombe, S. J. Stoyhoff, P. Davidson, & B. O'Sullivan (Eds.), *The Cambridge guide to language assessment* (pp. 133-139). Cambridge, UK: Cambridge University Press.

- Burger, M. 2008. "The alignment of teaching, learning and assessment in English language in grade 10 in district 9", Johannesburg (unpublished master thesis), University of South Africa.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). 2004. *Washback in language testing: Research contexts and methods*. London: LAWRENCE ERLBAUM ASSOCIATES, Publishers.
- Ching, P. 2008. A critical review of five language washback studies from 1995-2007: Methodological considerations. *JALT Testing & Evaluation SIG Newsletter*. 12 (2): 2-16.
- Cohen, L., Manion, L., & Morrison, K. 2007. *Research Methods in education (6<sup>th</sup> ed.)*. Oxon: Routledge.
- Creswell, J. 2009. *Research design (4<sup>th</sup> ed.)*. University of Nebraska-Lincoln: SAGE Publications, Inc.
- Cresswell, J. 2014. *Research design: Qualitative, quantitative and mixed methods approaches. (4<sup>th</sup> ed.)*. California: SAGE Publications Ltd.
- Creswell, J. W., & Plano Clark, V. L. 2011. *Designing and conducting mixed methods research (2<sup>nd</sup> ed.)*. London: SAGE Publications Ltd.
- Davies, A. 2008. *Assessing academic English: Testing English proficiency 1950–1989: The IELTS solution*. Cambridge, England: Cambridge University Press and Cambridge ESOL.
- Davies, A. 2012a. Ethical codes and unexpected consequences. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 455–68). London, England: Routledge.
- Davies, A. 2012b. Validity and soundness. *Language Testing*, 29 (1): 37–42.
- Davies, A. 2014. Fifty years of language assessment. In Antony Kunnan (Ed.), *The Companion to Language Assessment*, 11 (1): 1-18, John Wiley & Sons Inc, University of Edinburgh, Scotland. (<http://onlinelibrary.wiley.com>). (Accessed October 28, 2017).
- Djuric, M. 2008. Dealing with situations of positive and negative washback. *Scripta Manent*, 4 (1): 14-27.
- Five, H. & Nicole, D. 2013. Classroom test construction: The power of a table of specifications. *Practical Assessment, Research & Evaluation*, 18 (3): 67-81.
- Green, A. 2007. Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 7 (2): 75-97.
- Guma, A. 2010. *The Impact of test content validity on language teaching and learning*. Shaqra University: Shaqra University press.
- Hailu, D. 2015. *Assessing the content validity of teacher-made test in ELF class: The case of grade 11 at Ambo preparatory (MA thesis)*. Ambo University: Unpublished.
- Hawkey, R. 2006. *Impact theory and practice*. New York: Cambridge University Press.
- Hughes, A. 2003. *Testing for language teachers (2<sup>nd</sup> ed.)*. Cambridge: Cambridge University Press.

- Iyer, S. 2015. Impact of washback in English as second language classrooms – an investigation. *ELT Voices- International Journal for Teachers of English*, 5 (3): 88-91.
- Jabbarifar, T. 2009. The importance of classroom assessment and evaluation in educational system. *Proceedings of the 2<sup>nd</sup> International Conference of Teaching and Learning*, 2-4.
- Khan, I. S. 2013. Examination washback over learning and teaching. *Journal of Education and Practice*, 6 (1): 4-5.
- Kothari. 2004. *Research methodology* (2<sup>nd</sup> ed.). University of Rajasthan (India): New Age International Publisher .
- Motuma, H. 2014. An assessment of the students' reading ability in higher institutions: The case of Asela College of Teachers Education. *Journal of Science and Sustainable Development (JSSD)*, 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 (JSSD)*, 2015, 3 (1): 133-155.
- Motuma, H. 2017. Alignmet of primary school completion examination with the textbook: 2006-2009 in focus. Unpublished.
- Nigussie, T. 2002. Content validity of the ethiopian general secondary education certificate English examination (EGSECEE). Unplished *MA thesis*.
- Okeeffe, L. 2013. A framework for textbook analysis, *International Journal Review of Contemporary Learning Research*, 2 (1):1-13.
- Shannon, S. E. & Hsieh, H. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research*, 15 (9): 1277-1287.
- Shih, C. 2007. A new washback model of students' learning, *The Canadian Modern Language Review*, 64 (1): 135-161.
- Simachew, G. 2012. Washback of the university entrance English exam (UEEE) on teachers' and students' practices: The case of preparatory schools in Amhara national regional state. PhD Desertation, pp. 34-40.
- Sukyadi, D. & Mardiani, R. 2011. The washback effect of the English national examination (ENE) on English teachers' classroom teaching and students' learning. *Kata*, 13 (1): 96-111.
- Taylor, L. 2005. Washback and impact. *Key concepts in elt*, 154 .
- Thaidan, R. 2015. Washback in language testing. *International Journal of Science and Research*, 7.
- Weir, C. J. 2005. *Language testing and validation*. Palgrave Macmillan: Martin's Press.
- Wolming, S. & Wikstrom, C. 2010. The concept of validity in theory and practice: Assessment in Education, *Principles, Policy & Practice*, 17 (2): 117-132.
- Yasin, H. 2014. *Washback effet of grade ten national examination on teachers teaching practice* (Unpublished *Master's thesis*). Haramaya: Haramaya University.