Scientia
Publica
Media

🔓 **Research Article**

Assessing Final Exams for Minimum Learning Competency (MLC) Satisfaction in Seven Ethiopian Primary Schools

Workineh Kebede[1] ✉ iD

[1]Department of Curriculum Studies, Injibara College of Teacher Education, Ethiopia

Abstract

This study aimed to assess the quality of teacher-created final exams for Grade Four students in seven government primary schools in Injibara Town, Ethiopia, focusing on their alignment with Minimum Learning Competency (MLC) standards. A qualitative content analysis was conducted on 505 exam items, with two independent raters evaluating their adherence to MLC criteria. The inter-rater reliability, measured using Cohen's Kappa coefficient, was deemed acceptable. Additionally, a survey was administered to teachers to gauge their awareness of MLC and identify challenges in exam design and administration. Both quantitative and qualitative methods were employed to analyze the data. The results indicated that a substantial portion of the exam items did not meet MLC standards, often failing to evaluate the full spectrum of competencies outlined in the MLC framework. Most questions focused on lower-order cognitive skills, such as knowledge and comprehension, while higher-order cognitive skills and non-cognitive domains like affective and psychomotor competencies were largely overlooked. Open-ended survey responses revealed a lack of clarity among teachers in designing assessments aligned with MLC, insufficient access to resources (e.g., teaching materials, textbooks, sample exams, and digital tools), heavy workloads, and administrative challenges such as tight deadlines, lack of departmental coordination, and insufficient support from school leadership—all of which hindered the design and implementation of competency-based assessments. To address these challenges, the study recommends introducing a Competency-Based Assessment (CBA) course in teacher education programs, providing targeted training on MLC-aligned exam development, recruiting qualified teacher educators, ensuring adequate resources, reducing teacher workloads, strengthening school leadership, and establishing an exam evaluation committee. Further research in this area should also be conducted. These interventions aim to improve assessment quality, enhance teachers' capacity to design competency-based exams, and advance the implementation of MLC standards in Ethiopian primary schools.

Keywords: Bloom's Taxonomy, Competency-Based Assessment (CBA), Ethiopia, Grade Four, Minimum Learning Competency (MLC), Written Final Exams

1. INTRODUCTION

According to Idris et al. (2012), education plays a crucial role in fostering economic growth, societal development, and the cultivation of knowledge and skills among the population, particularly in shaping the character of young individuals. Ozturk (2001) emphasizes that education is a cornerstone of development, arguing that sustainable economic growth is unattainable without significant investment in human capital.

UNESCO (2001) emphasizes that individuals in society possess varying levels of ability, indicating that not everyone reaches the same level of competency. A key goal of education is to cultivate individuals into responsible citizens, which requires the development of specific competencies. When designing educational programs, it is essential to define the desired level of competency, with teachers playing a central

role in guiding students toward achieving these goals. Competency is built through a combination of knowledge, skills, and attitudes, which are developed through consistent effort, practice, and perseverance. UNESCO (2001) further explains that while people may acquire different types of competencies—such as educational or physical—these are achieved through sustained effort and practice. Competency is not innate but rather a combination of acquired knowledge, practical skills, and the right mindset, all of which are essential for personal and societal growth.

Competencies can originate from various sources, often emerging from program outcomes when institutions transition from traditional, time-based instruction to a competency-based learning model. These competencies are grounded in the essential knowledge and skills required for success (Baughman et al., 2012). As Carraccio et al. (2002) note, while outcomes focus on the acquisition of knowledge, competencies prioritize the practical application of that knowledge. In the context of general education, these are often termed institutional competencies and may encompass skills such as public speaking, critical thinking, effective communication, ethical reasoning, cultural understanding, and the ability to integrate and apply information (Cecil & Krohn, 2012).

Drisko (2014) emphasizes the importance of making competencies measurable and highlights the critical role of effective assessment. To accurately gauge competencies, a variety of assessment methods must be utilized to determine whether a student has truly mastered them. Each competency should be assessed multiple times and through diverse approaches, such as multiple-choice tests, essays, presentations, and performance-based tasks. Competency assessments should not focus solely on individual skills, but rather on how multiple competencies combine in real-world scenarios. Additionally, Roy (2016) stresses that competencies should be both observable and measurable, enabling clear evaluation through a defined process.

Assessments intended to measure competencies should go beyond just assessing knowledge and skills; they must also evaluate a student's ability to integrate, synthesize, and apply that knowledge within a community of practice. The assessments should be sensitive enough to detect growth and development in learning over time, meaning that competencies should be evaluated multiple times throughout the course. Properly designed assessments are essential to ensure that they accurately measure the intended competencies (Baughman et al., 2012; Drisko, 2014).

The Ministry of Education of Ethiopia (MoE) (2010) stresses the need for clearly defined learner competencies at all educational levels, with a structured learning progression that aligns with international standards and takes students' cognitive development into account. Competencies should be outlined in terms of knowledge, comprehension, and skills. Similarly, the Singapore Workforce Development Agency (SWDA) (2012) describes competency-based assessments as a means of collecting evidence that demonstrates a learner's ability to meet specific standards. Such assessments are based on clearly defined outcomes within competency standards, and they should allow assessors to determine whether those outcomes have been achieved. Furthermore, these assessments must ensure that the evidence gathered clearly shows a learner has met each performance criterion.

In the 1970s, a shift occurred from traditional content-based teaching to focusing on learning outcomes, which brought the issue of assessment to the forefront. Rather than structuring curricula around what the teacher needs to teach, the concept of teaching by objectives suggested designing curricula around what students should be able to master, with a focus on clearly defined, operational objectives. Today, most education systems worldwide have embraced the idea of placing competencies at the center of curricula (Roegiers, 2016).

Competency-based assessment and traditional norm-referenced psychometric assessments differ significantly in how they evaluate results. While competency-based assessment measures an individual's performance against a set standard, norm-referenced assessment compares a person's performance to that of others (Hager, Gonczi & Athanasou, 2006). Competency-based assessment is aimed at gathering objective and valuable information about students' learning to guide decision-making and provide insights into why certain training experiences are successful or not (Delgado et al., 2005, as cited in Munoz & Araya, 2017).

Cano (2008), as referenced by Munoz and Araya (2017), views competency-based assessment as a key component of a new assessment approach. First, it is seen as a pivotal point in the learning process because it is both influenced by and a driver of learning. Competency-based assessment can guide curriculum development, foster spaces for innovative formative learning, and set clear evaluation standards. Secondly, it supports the use of diverse assessment tools in a multi-dimensional way and encourages the involvement of various educational stakeholders, such as teachers, students, and administrators. Both definitions, according to Munoz and Araya, suggest that competency-based assessment is a comprehensive framework that should be carefully designed to enhance the evaluation process and effectively showcase the progress made by students, educators, and educational institutions.

Competency-based assessment should support students in becoming more self-aware of their learning and enhance their ability to regulate their own educational progress. Jaimes and Callejas (2009), as cited in Munoz and Araya (2017), emphasize that assessment should address the three key dimensions of learning: knowing, knowing how to do, and knowing how to be, all from the perspective of the learning goals. According to Munoz and Araya (2017), competency-based evaluation can positively impact both learning outcomes and teaching practices. Thomas and Cross (2010) also highlight that effective assessment begins with clear goals, which requires thoughtful planning by educators. Teachers must carefully consider which competencies are essential to teach within the course content and what students need to learn.

Various competency-based assessment methods, as identified by Schultz and Schultz (1990) and Meyer (1996), include approaches such as case studies, in-basket exercises, job sample or skill tests, direct observation, oral and written exams, portfolios, and projects and assignments. Exams and evaluations are crucial elements of the educational process, acting as a scientific means of confirming performance. In education, final exams and recognition of achievements are vital in preparing students for the labor market, making it essential that exams are designed to reflect competencies. Golubina and Loser (2018) state that the primary goal of educational institutions is to enhance the quality of their offerings, and one effective way to ensure high standards is through examinations that align with established criteria.

Competency-based assessment and traditional norm-referenced psychometric assessments differ significantly in how they evaluate performance. While competency-based assessment measures an individual's performance against a predefined standard, norm-referenced assessment compare a person's performance to that of others (Hager et al., 2006). Competency-based assessment is a process designed to gather objective and useful information about students' learning, which can then inform decision-making and provide insights into the factors behind the success or failure of the training (Delgado et al., 2005, as cited in Munoz & Araya, 2017).

Bloom's taxonomy outlines three primary learning domains: cognitive, affective, and psychomotor, a framework developed by a team of researchers led by Benjamin Bloom. The cognitive domain focuses on mental processes, specifically those related to thinking, and is used to assess a student's intellectual abilities through exams. This domain addresses objectives related to knowledge acquisition and the development of critical thinking skills. It is structured into six levels, progressing from basic to more complex, including knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom, 1956).

The affective domain encompasses objectives related to changes in emotions, feelings, interests, values, attitudes, and the development of appreciations and appropriate adjustments. In contrast, the psychomotor domain focuses on manipulative or motor skills, as well as mechanical or physical knowledge. The affective domain is structured into five key categories: receiving, responding, valuing, organization, and characterization. Similarly, the psychomotor domain, as outlined by Dave (1970), includes five levels: imitation, manipulation, precision, articulation, and naturalization. Simpson (1972) expanded this framework by proposing a more detailed hierarchy with seven levels: perception, set (mindset), guided response, mechanism, complex overt response, adaptation, and origination. Both domains organize their categories or levels in a progression from simple to complex and from concrete to abstract concepts.

In the cognitive domain, objectives that focus solely on recalling or recognizing information are classified as knowledge and comprehension. On the other hand, objectives that require a deeper understanding and application of knowledge fall within the categories ranging from comprehension to synthesis, which are considered the most critical educational goals. This classification forms a foundation for shifting curricula and exams toward more complex objectives (Krathwohl, 2002).

The cognitive levels of questions typically posed by teachers in exams tend to remain within the lower levels of learning—knowledge, comprehension, and application. As a result, there is less emphasis on evaluating higher-order cognitive skills such as analysis, synthesis, and evaluation (Anees, 2017). Alhashem and Agha (2020) argue that a comprehensive approach to final exams, which includes all three domains of learning, is essential for effective student assessment.

A tracer study by Abrham et al. (2011) found that diploma graduates from Amhara Region's colleges of teacher education displayed low levels of competence in areas such as professional skills, lesson planning, student learning assessment, facilitating student learning, engaging in professional development, and mastering educational training programs and curricula. The study also highlighted the need to avoid assessing student competence through group assignments, advocating for a more individualized assessment approach that aligns with the instructional process.

Abrham et al. (2011) and Workineh et al. (2014) found that the professional competence of primary school teachers who graduated from colleges of teacher education in the Amhara Region did not significantly differ based on gender (male or female), teaching experience (novice or experienced), training modality (cluster or linear), training college, or training program (regular, extension, or summer). However, Abrham et al. (2011) noted variations in teaching competencies among graduates from different academic departments. Specifically, graduates in language, aesthetics, and physical education demonstrated lower levels of competence, while those in natural sciences, mathematics, and social sciences exhibited moderate levels of competency.

The professional expertise of secondary school teachers, as defined by the Ministry of Education (MoE) in 2003, falls short of the established benchmarks and expectations for their roles. Additionally, research conducted by Asnakew and Lake (2015) highlighted that secondary school teachers in the Amhara National Regional State struggled to exhibit proficiency in evaluating student learning. Their study also noted that assessment competence did not vary significantly based on gender but tended to improve with teaching experience before eventually declining. Although not statistically significant, language teachers were observed to be slightly more skilled in assessing student learning compared to teachers of other subjects. Asnakew and Lake further emphasized that teachers lack the necessary knowledge and skills to effectively evaluate student achievement, which could undermine the quality of education in the Amhara Region, exacerbating its already substandard educational standards. These findings underscore that student assessment remains a critical challenge in both primary and secondary schools across the region.

Lenburg (2007) emphasized that the ability of teacher educators to gauge teaching effectiveness is closely tied to teachers' capacity to assess student learning. A fundamental principle of effective curriculum design and teaching is ensuring alignment between learning outcomes, instructional methods, and assessment strategies. Designing appropriate assessments is only possible when clear standards for achievement have been established.

Abrham et al. (2011) highlighted that diploma graduates from Amhara Region Colleges of Teachers' Education, who were teaching in both the first and second cycles of primary schools, demonstrated limited competence in assessing student learning. This deficiency was consistent across gender, teaching experience, and subject specialization.

The notion of competence has deep historical roots, dating back approximately 3,000 years to ancient China, where written civil service exams were used to select candidates for government positions (Hoge et al., 2005). Horton (2000) noted that during the medieval period, the apprenticeship system emerged, where apprentices learned skills under the guidance of a master and received credentials upon meeting the required standards of craftsmanship.

The industrial revolution brought significant socioeconomic transformations across agriculture, manufacturing, production, and transportation sectors. This period also saw the emergence of studies focusing on work, job roles, and the skills necessary to perform them (Horton, 2000). By the late 19th and early 20th centuries, the concept of social efficiency gained prominence in the United States. Frederick Taylor, often regarded as the father of scientific management, played a pivotal role in shaping management theory and practice. Taylor (1911) advocated for a more detailed division of labor, simplified job tasks, increased managerial oversight, and the use of systematic time and motion studies for cost accounting. These developments were closely linked to the rise of the competence concept.

The concept of competence evolved further within management practices, particularly in the context of work and employee selection (Hoge et al., 2005). During the 1930s in the United States, the Roosevelt administration emphasized the functional analysis of jobs, leading to the creation of a dictionary of occupational titles that outlined the knowledge and skills associated with various professions (Horton, 2000). In the 1940s and 1950s, researchers began systematically identifying and analyzing broad performance factors (Flanagan, 1954; Fleishman, 1953). By the 1960s, psychologists focused on identifying individual variables that could predict job performance effectively while minimizing bias against specific subgroups (Shippmann et al., 2000).

In 1978, the U.S. government introduced the Uniform Guidelines on Employee Selection Procedures, which mandated that employee selection be based on job-related qualifications derived from analyzing essential work behaviors and desired outcomes (Shippmann et al., 2000). Later, in 1994, Congress established the National Skill Standards Board (NSSB), which became the National Skill Standards Board Institute (NSSBI) in 2003. This body aimed to develop a voluntary national system of skill standards, assessments, and certifications to strengthen the United States' competitiveness in the global economy (Horton, 2000). Standards were defined as benchmarks that specify the work to be performed, the quality of performance, and the level of knowledge and skills required (McCain, 2002).

The competence movement not only promoted efficiency in industrial and business sectors but also reshaped education through the rise of competency-based education, an approach rooted in industrial and business models that emphasized defining outcomes as measurable behavioral objectives (Tuxworth, 1989). Widely believed to have originated in teacher education, this approach gained momentum during the First World War when educators were tasked with training skilled tradesmen and technicians to instruct inexperienced military workers, highlighting the importance of structured, outcome-focused teaching methods (Hoffmann, 1999; Neumann, 1979; Tuxworth, 1989).

While there is no universally agreed-upon definition of competence, various researchers and organizations have attempted to define it. Generally, competence is associated with the knowledge, skills, and attitudes required performing the activities of a specific occupation or function effectively, meeting the expected standards (Bassellier et al., 2001). For the purposes of this research, the definition provided by Bassellier et al. (2001) is adopted. Similarly, the National Postsecondary Education Cooperative (NPEC) (2017) defines competence as a blend of skills, abilities, and knowledge necessary to execute a specific task within a given context.

The primary aim of this research was to investigate whether the Grade Four subject teachers' final written exams in seven Ethiopian general primary schools—Kossober, Injibara, Bata, Anqish, Bahunk, Andinet, and Hintset—aligned with the established minimum learning competency (MLC). To achieve this, the study sought to address the following key research questions:

1. To what extent do Grade Four subject final exams meet the specified minimum learning competency (MLC)?

2. Is there a significant difference in how Grade Four final exam items align with MLC across different subjects?

3. How do Grade Four final exams reflect the three domains of learning—cognitive, affective, and psychomotor—and their respective levels?

4. What challenges do Grade Four teachers encounter when developing written exams?

2. METHODS AND PROCEDURES

The main objective of this study was to investigate the extent to which Grade Four teacher-created written final exams in seven government primary schools in Injibara Town, Ethiopia, meet the defined minimum learning competencies (MLC).

2.1. Research Method

To accomplish this goal, the study employed qualitative content analysis as its primary research method. Parveen and Showkat (2017) describe content analysis as a technique for examining the contexts, meanings, and intentions within messages. They also emphasize that content analysis is cost-effective, adaptable to both qualitative and quantitative research, and remains unbiased by the research process itself. In this study, conceptual content analysis was utilized. As outlined by Devi (2009), conceptual content analysis involves identifying and quantifying the presence of specific words, phrases, concepts, or themes in a text, allowing researchers to draw conclusions based on emerging patterns. This method focuses on detecting and measuring the frequency of key concepts. Parveen and Showkat (2017) further note that conceptual content analysis integrates both qualitative and quantitative data. Creswell (2014) supports this approach, arguing that combining quantitative and qualitative methods offers a more thorough understanding of complex issues and phenomena than using either method in isolation.

2.2. Population, Sample, and Sampling Techniques

The study focused on Grade Four teacher-created final exams across all five subjects—Amharic, English, Mathematics, Environmental Science, and Aesthetics—from seven government primary schools in Injibara Town, located in the Awi Zone of Ethiopia. Notably, the exams for Mathematics, Environmental Science, Aesthetics, and Amharic itself were administered in the Amharic language.

The seven schools within Injibara Administrative Town were intentionally chosen for the research due to their accessibility and convenience for the researcher. All Grade Four teachers from these schools participated in the survey, employing a comprehensive or total sampling approach. As Shannon and Bradshaw (2002) highlight, including the entire population in a study enhances the representativeness of the findings, encourages broader participation, minimizes accuracy concerns, and streamlines the administration process, as every individual receives a survey. The analysis was limited to the teacher-made final exams for Grade Four subjects, as this provided a practical and manageable dataset for the research.

2.3. Data Collection Instruments

The data collection instruments for this study consisted of document analysis and a questionnaire. Document analysis was employed to evaluate the Grade Four teacher-made written final exams across all subjects, with a focus on determining whether they aligned with the MLC standards. This process involved systematically categorizing each exam item into three learning domains: cognitive, affective, and psychomotor. Document analysis is a structured approach to reviewing documents, enabling researchers to extract meaning, develop understanding, and generate empirical knowledge (Corbin & Strauss, 2008, as cited in Bowen, 2009).

Additionally, a questionnaire comprising both closed- and open-ended questions was distributed to all Grade Four teachers at the participating schools. The questionnaire aimed to gauge the teachers' awareness of the MLC and identify the challenges they encounter in preparing exams, with the ultimate goal of enhancing the quality of teacher-made written assessments. As Carter and Williamson (1996) noted, designing a questionnaire in a standardized manner offers a quick and efficient means of data collection. It facilitates the gathering of information from a large group, and the responses tend to be more objective or less subjective in nature.

2.4. Data Analysis Techniques

In this study, a mixed-methods approach was utilized to analyze the data gathered through content analysis. Quantitative data were examined using descriptive statistics, such as frequency counts and percentages or proportions, as well as inferential statistics, including confidence intervals, chi-square tests, and standardized residual tests. Qualitative methods were also employed to provide deeper insights and context to the findings. This combination of approaches allowed for a comprehensive understanding of the data, balancing numerical analysis with nuanced interpretation.

For the qualitative data, a categorical method was used to analyze the exam items, identifying those that met or did not meet the minimum learning competency (MLC) standards. These items were categorized according to their respective learning domains. Additionally, thematic analysis was applied to the open-ended responses from the teachers' questionnaires. The responses were analyzed qualitatively, with findings presented through explanation, narration, and quotation where necessary.

2.5. Exam Analysis Techniques and Procedures

The evaluation of the exams to assess their adequacy in meeting the predetermined MLC standards and their categorization into appropriate learning domains and sub-domains was carried out by two teacher educators at Injibara College of Teacher Education, Ethiopia. One rater held an M.Ed. in Curriculum Studies, while the other held an M.Ed. in Educational Psychology. The raters performed their evaluations independently, after being thoroughly briefed on the study's objectives and provided with the syllabus for all five Grade Four subjects, which included the MLC standards.

Cohen's Kappa coefficient ($\varkappa$) was calculated to measure inter-rater reliability, resulting in a value of 0.56, with an agreement percentage of 78.02%. As outlined by Landis and Koch (1977), a Kappa value of 0.56 signifies moderate agreement between the two raters. This indicates a reasonable degree of consistency in the evaluations provided by the two assessors.

The data collection process for the final exams followed a structured approach to ensure alignment with the study's objectives. Initially, the two raters were briefed on the study's aims, provided with the MLC standards for each subject, and introduced to the three learning domains—cognitive, affective, and psychomotor—along with their corresponding levels of learning. Each exam item across all subjects was systematically coded for analysis. The raters then independently assessed whether the items met the MLC standards specified in the syllabus, classifying each item according to the appropriate learning domain and level. Following this, the researcher collected the ratings from both raters, computed Cohen's Kappa coefficient to evaluate inter-rater reliability, and analyzed the data to derive conclusions consistent with the study's goals.

Table 1. Subjects' Item Analysis Results for Satisfying MLC and Kappa Coefficient (k)

Raters' decision on the number of items	Subjects					Total	Percent
	Amharic	English	Math	Environmental Science	Aesthetics		
Both raters agreed that the number of item that satisfied MLC	36	28	65	44	16	189	37.4
Both raters agreed that the number of items that didn't satisfy MLC	40	49	25	42	49	205	40.6
Only the first rater (rater A) agreed that the number of items that satisfied MLC	9	15	13	14	9	60	11.9
Only the second rater (rater B) agreed that the number of items that satisfied MLC	10	4	11	15	11	51	10.1
Total	95	96	114	115	85	505	100.0

(Note: "62.6" appears in the Percent column aligned with the "Only the second rater" row group.)

3. RESULTS

This section provides an analysis and interpretation of the teacher-made final exam items, focusing on their alignment with the Minimum Learning Competencies (MLC) standards. The analysis seeks to determine whether there is a statistically significant difference between the number of exam items that meet the MLC criteria and those that do not. Additionally, it explores potential variations in the distribution of exam items across the three primary learning domains—cognitive, affective, and psychomotor—as well as their respective sub-domains or levels. Through this evaluation, the study aims to assess the extent to which the exams adhere to the prescribed standards and how well they address the diverse dimensions of learning.

The section also assesses the extent to which the exam items align with the specific MLC standards for each domain and level. Additionally, it identifies the challenges encountered by teachers during the construction of written final exams, providing insights into the difficulties they face in ensuring that their assessments adequately reflect the intended learning outcomes. The findings from this analysis are crucial for understanding how well the exams support the intended educational standards and for identifying areas for improvement in the exam design process.

Table 1 shows the exam item analysis of the five subjects for satisfying MLC as rated by two raters. Only 189 (37.4%) of the final exams of the survey schools out of a total of 505 items were labeled by both raters as satisfying MLC. However, a greater number of the exam items (62.6%) were labeled as not meeting the MLC standard. The inter-rater reliability, measured using Cohen's Kappa coefficient (k), was found to be 0.56, with an agreement percentage of 78.02%. Based on the interpretation by Landis and Koch (1977), this indicates a moderate level of agreement.

Table 2. Confidence Intervals (CI) for Exam Items that Satisfy and Do Not Satisfy the MLC

Raters' decision on the number of items	Confidence Interval (CI) at 95% CI & $Z_{\alpha/2}$=1.96
Confidence interval (CI) for items that satisfied MLC ; for 189 items	[33.21%, 41.64%]
Confidence interval for items that didn't satisfy MLC ; for 316 items	[58.35%, 66.78%]

As shown in Table 2, there is a 95% probability that the grade four teacher-made written final exams, which are expected to meet the stated MLC standards, fall within a range of 33.21% to 41.64%. This indicates that only a small proportion of the exams can be considered to satisfy the MLC. In contrast, a large proportion of exam items, with a 95% confidence interval between 58.35% and 66.78%, do not meet the MLC standards. Since these two confidence intervals do not overlap at the 95% confidence level, it suggests a statistically significant difference between the number of items that satisfy the MLC and those that do not. This, in turn, indicates that the majority of the exam items fail to meet the MLC standard, rendering them invalid in terms of alignment with the MLC.

Table 3. Chi-square Computation by Subjects for Exam Items Satisfying MLC

Subjects	Observed (O)	Expected (E)	(O-E)	$R=\frac{O-E}{\sqrt{E}}$	
Amharic	36	35.6	0.4	0.067	
English	28	35.9	-7.9	-1.318	χ^2calculated=
Mathematics	65	42.7	22.3	3.412	21.2625
Environmental science	44	43.0	1	0.152	
Aesthetics	16	31.8	-15.8	-2.8	
Total	189	189	-		

Note: χ^2critical = 9.488, df = 4, α = 0.05

In Table 3, the expected frequencies were calculated based on the number of items analyzed for each subject, taking into account the 189 items labeled by both raters as satisfying MLC out of the total 505 questions or items analyzed. Simple proportions derived from the totals were used to determine the expected number of items addressed in Table 3. These expected frequencies were then compared with the actual (observed) frequencies to compute the chi-square statistic for comparison. The expected frequencies were calculated as follows:

[Amharic:(95/505)*189=35.6;English:(96/505)*189=35.9;Mathematics:(114/505)*189=42.7;Environmental Science:(115/505)*189=43.0 & Aesthetics:(85/505)*189=31.8]. Since Chi-square critical = 9.488 exceeds the Chi-square calculated 2=21.2625, there was a statistically significant evidence at α=0.05 that the observed frequencies and expected frequencies were not the same. Hence, there was statistically significant difference between the observed and expected number of items across subjects satisfying MLC.

According to the rule of thumb, a statistically significant difference between expected and observed frequencies exists if the standardized residual (R) (calculated as the difference between observed and expected frequencies divided by the square root of the expected frequency) exceeds 2 or falls below -2. Applying this criterion, a statistically significant difference was observed between the observed and expected number of items satisfying the MLC for Mathematics (R = 3.412) and Aesthetics (R = 2.8). This indicates that the teacher- made Mathematics final exams were better aligned with the MLC standards compared to other subjects in this study, as evidenced by the positive R-value greater than 2. Conversely, the Aesthetics final exam was found to be the least aligned with the MLC standards among the subjects analyzed.

Besides, in Table 3, for English, Amharic, and Environmental Science, no statistically significant differences were observed between the observed and expected frequencies. However, the exam items for English were slightly fewer in meeting the MLC standards compared to Amharic and Environmental Science, as indicated by the negative R-value. Overall, the standardized residuals for Aesthetics and Mathematics were the primary contributors to the total chi-square value.

Table 4. Classification of Items Satisfying MLC by Domains and Levels of Learning

| Subjects | Domains of Learning | | | | | Total | Percent |
| | Cognitive | | | Affective | Psychomotor | | |
	Knowledge	Comprehension	Application				
Amharic	2	18	-	-	-	20	24.7
English	3	6	-	-	-	9	11.1
Mathematics	11	3	7	-	-	21	25.9
Environmental science	16	9	-	-	-	25	30.9
Aesthetics	2	1	3	-	-	6	7.4
Total	34	37	10	-	-	81	100.0
Percent	42.0	45.7	12.3	-	-	100.0	

Table 4 shows that all 505 exam items were labeled as cognitive-type questions, with affective and psychomotor-type questions completely excluded. Of the 189 items labeled as satisfying MLC, only 81 (42.8%) were assigned the same sub-domains/levels of learning by both raters. In contrast, a larger portion (57.2%) of the MLC-satisfying items was categorized differently by the raters with respect to their levels of learning. A significant proportion of the exam items focused on recall and comprehension, with only a slight difference between the two. Together, recall and comprehension questions accounted for approximately 87.7% of the total, leaving little room for other levels of cognitive learning.

Table 5 shows that approximately 93.9% of respondents reported being sufficiently aware of Minimum Learning Competency (MLC), while only 6.1% indicated a lack of awareness. This addresses the question, "Are you sufficiently aware of MLC?", as the majority of teachers confirmed their awareness. Additionally, 100% of respondents stated that they took MLC into account when preparing or setting grade four final exams, adhering to the established MLC guidelines. This aligns with the question, "Do you take MLC into account when preparing final exams?", as all teachers claimed to incorporate MLC into their exam preparation.

However, a notable contradiction emerges: despite 6.1% of teachers reporting insufficient knowledge of MLC, they still claimed to incorporate it when setting final exams. This discrepancy highlights a gap between teachers' self-reported understanding of MLC and its actual implementation in exam preparation, as evidenced by the item analysis, which revealed that the majority of exam items did not meet the MLC standard.

Table 5. Teachers' Awareness about MLC and Competency-Based Exam

Item	Response				Total	Confidence Interval (CI) of responses at 95% confidence level, $Z_{\alpha/2}$=1.96 Response	
	Yes		No			Yes	No
	Count	Percent	Count	Percent			
Are you sufficiently aware of MLC?	46	93.9	3	6.1	100.0	[87.16%, 100%]	[2.69%, 9.54%]
Do you take MLC into account when preparing final exams?	49	100.0	-	-	100.0	100%	-

The primary objective of this study was to assess the alignment of Grade Four teacher-made written final exams with the specified Minimum Learning Competencies (MLC) in seven primary schools in Injibara town, Ethiopia. The analysis revealed that only 37.4% of the exam items (189 out of 505) met the MLC criteria, while 62.6% failed to satisfy the established standards, highlighting a significant gap in the quality of assessments. In addition, all exam items were of a cognitive nature, with both the affective and psychomotor domains entirely neglected. The exam items primarily assessed lower-level cognitive skills, such as recall and comprehension, rather than higher-order thinking skills. Furthermore, there was considerable variation in how exam items met the MLC across subjects: while Mathematics exams showed a relatively better alignment, Aesthetics exams were found to be highly deficient in meeting the MLC criteria. The study also revealed significant challenges in the exam development process. Teachers displayed a limited understanding of the MLC, lacked sufficient teaching materials, and faced administrative challenges in implementing competency-based assessments effectively. These findings underscore the need for substantial improvements in both teacher training and assessment design to ensure exams more accurately reflect students' learning outcomes.

The Cohen's Kappa coefficient (k = 0.56) and the percentage of agreement (78.02%) indicated a moderate level of inter-rater reliability, affirming the validity of the instrument used for the analysis. The 95% confidence interval confirmed a statistically significant difference between the number of MLC-compliant and non-compliant items, further emphasizing the inadequacy of the assessments in meeting the standards. This result stands in contrast to the findings of Golubina and Loser (2018), who argued that adherence to predetermined assessment standards is crucial for ensuring the quality of education.

A striking pattern emerged from the content analysis, as all 505 exam items were found to focus solely on the cognitive domain of learning. Not a single item assessed the affective or psychomotor domains, which are equally essential for providing a well-rounded educational experience. Among the items that did meet the Minimum Learning Competency (MLC) criteria, 87.7% were confined to the lower levels of the cognitive domain, primarily addressing knowledge and comprehension. This overemphasis on basic cognitive skills, at the expense of higher-order cognitive skills such as analysis, synthesis, and evaluation, suggests a narrow approach to assessing students' overall learning. These findings align with the work of Alhashem and Agha (2020), who argue for a more holistic approach to assessments that includes all three learning domains—cognitive, affective, and psychomotor. Similarly, Nessipbayeva (2012) emphasized that effective competency assessments should encompass not only knowledge but also skills and attitudes, which were notably absent in the exams analyzed here. In agreement with these scholars, Rovai et al. (2009) asserted that perceived learning should be measured comprehensively, utilizing all three domains to fully capture students' competencies.

These results align with Anees (2017), who observed a lack of higher-order questions in assessments, and are consistent with Haymanot (2018), who stressed the need for exam items that encourage problem-solving and critical thinking. Moreover, the findings echoed Hughes (1996), who identified common challenges in exam preparation, such as ambiguous items, unclear instructions, and poorly designed test layouts, all of which can undermine the quality of assessments. Weir (1995) also outlined essential characteristics of well-designed exams, including appropriate difficulty levels, effective discrimination between student performances, and clear and unbiased items—all of which were often lacking in the analyzed exams.

Furthermore, the study revealed significant variations in the alignment of exam items with MLC across subjects. Mathematics exams demonstrated a statistically significant positive residual (R = 3.412), indicating a stronger alignment with MLC standards. In contrast, Aesthetics exams showed a negative residual (R = -2.8), reflecting a weaker alignment. This pattern aligns with findings from Abrham et al. (2011) and Workineh et al. (2014), who noted that primary school teachers in the Amhara Region, particularly in subjects like Language and Aesthetics, often lack sufficient competencies in assessing student learning. These findings highlight a critical gap, emphasizing the need for targeted professional development in assessment practices, especially in subjects traditionally viewed as less technical. Addressing these disparities is essential to ensure consistent alignment with MLC standards across all subjects.

The inadequate performance of Aesthetics teachers in assessment could be attributed to several factors, including the insufficient preparation and training of teacher educators in this field. As noted by Lenburg (2007), the ability of teachers to assess student learning is often shaped by the competence of their teacher educators. In the Amhara Region, many teacher educators in subjects like Music, Art, and Health and Physical Education (HPE) hold only diplomas and may lack the necessary expertise to provide high-quality training in assessment. This suggests a downstream impact from teacher training institutions to primary schools, where teachers may lack the skills to design valid assessments.

One of the more concerning findings of this research was the disconnect between teachers' reported practices and their actual understanding of the MLC. While 100% of the teachers claimed to consider the MLC when developing exams, 6% of the respondents admitted to having no clear understanding of the MLC. This contradiction suggests that teachers may lack a deep understanding of what the MLC entails and how to effectively implement it in exam design. This issue mirrors the challenges identified by Johnston (2015), who pointed to common problems in exam development, such as overly simplistic questions that measure only factual recall, ambiguous wording, and a failure to align exams with course goals and competencies.

The analysis of qualitative data from teachers' open-ended responses revealed four major challenges in exam setting: First, teachers demonstrated a lack of knowledge and understanding regarding how to design exams that align with the Minimum Learning Competencies (MLC). Second, there was a limited availability of material resources needed to create high-quality exams, hindering effective assessment development. Third, many teachers exhibited an unwillingness or lack of readiness to invest the necessary time and effort in preparing exams, which could affect the quality of the assessments. Finally, teachers faced significant administrative challenges, such as insufficient support or guidance in the exam preparation process, which further complicated the creation of well-structured and valid exams.

These challenges are consistent with findings from Oluoch (2014), who highlighted issues such as poorly structured questions, excessive overlap of content, and ambiguous instructions as major obstacles to effective assessment. Similarly, Usman (2016) emphasized the importance of school leadership in ensuring that adequate resources are available for teaching and assessment, pointing to a systemic issue in the broader educational infrastructure that hampers effective assessment practices.

The findings underline the critical need for policy-level support to drive educational reforms in assessment practices. Policymakers should prioritize revising national assessment frameworks to ensure exams comprehensively evaluate students' competencies across all learning domains. Additionally, teacher education institutions must be empowered to equip educators with the skills necessary to design balanced and effective assessments. Furthermore, research into alternative assessment methods, such as project-based learning and performance assessments, could offer valuable insights into more effectively evaluating competencies, particularly in the affective and psychomotor domains. This study lays the groundwork for broader investigations into how educational systems can evolve to deliver more holistic, competency-based assessments that better address the diverse needs of students.

In conclusion, this study underscores the pressing need for comprehensive reforms in both teacher training and assessment practices in Ethiopian primary schools. Teachers must receive clearer guidance on the MLC, along with professional development in designing assessments that address all three learning domains—cognitive, affective, and psychomotor. Furthermore, addressing material and administrative challenges will be crucial in supporting teachers to create high-quality exams that accurately reflect student learning outcomes. These efforts will contribute to the improvement of the overall quality of education and

help ensure that assessments truly measure the full range of competencies that students are expected to develop.

4. DISCUSSION

The primary objective of this study was to assess the alignment of Grade Four teacher-made written final exams with the specified Minimum Learning Competencies (MLC) in seven primary schools in Injibara town, Ethiopia. The analysis revealed that only 37.4% of the exam items (189 out of 505) met the MLC criteria, while 62.6% failed to satisfy the established standards, highlighting a significant gap in the quality of assessments. In addition, all exam items were of a cognitive nature, with both the affective and psychomotor domains entirely neglected. The exam items primarily assessed lower-level cognitive skills, such as recall and comprehension, rather than higher-order thinking skills. Furthermore, there was considerable variation in how exam items met the MLC across subjects: while Mathematics exams showed a relatively better alignment, Aesthetics exams were found to be highly deficient in meeting the MLC criteria. The study also revealed significant challenges in the exam development process. Teachers displayed a limited understanding of the MLC, lacked sufficient teaching materials, and faced administrative challenges in implementing competency-based assessments effectively. These findings underscore the need for substantial improvements in both teacher training and assessment design to ensure exams more accurately reflect students' learning outcomes.

The Cohen's Kappa coefficient (k = 0.56) and the percentage of agreement (78.02%) indicated a moderate level of inter-rater reliability, affirming the validity of the instrument used for the analysis. The 95% confidence interval confirmed a statistically significant difference between the number of MLC-compliant and non-compliant items, further emphasizing the inadequacy of the assessments in meeting the standards. This result stands in contrast to the findings of Golubina and Loser (2018), who argued that adherence to predetermined assessment standards is crucial for ensuring the quality of education.

A striking pattern emerged from the content analysis, as all 505 exam items were found to focus solely on the cognitive domain of learning. Not a single item assessed the affective or psychomotor domains, which are equally essential for providing a well-rounded educational experience. Among the items that did meet the Minimum Learning Competency (MLC) criteria, 87.7% were confined to the lower levels of the cognitive domain, primarily addressing knowledge and comprehension. This overemphasis on basic cognitive skills, at the expense of higher-order cognitive skills such as analysis, synthesis, and evaluation, suggests a narrow approach to assessing students' overall learning. These findings align with the work of Alhashem and Agha (2020), who argue for a more holistic approach to assessments that includes all three learning domains—cognitive, affective, and psychomotor. Similarly, Nessipbayeva (2012) emphasized that effective competency assessments should encompass not only knowledge but also skills and attitudes, which were notably absent in the exams analyzed here. In agreement with these scholars, Rovai et al. (2009) asserted that perceived learning should be measured comprehensively, utilizing all three domains to fully capture students' competencies.

These results align with Anees (2017), who observed a lack of higher-order questions in assessments, and are consistent with Haymanot (2018), who stressed the need for exam items that encourage problem-solving and critical thinking. Moreover, the findings echoed Hughes (1996), who identified common challenges in exam preparation, such as ambiguous items, unclear instructions, and poorly designed test layouts, all of which can undermine the quality of assessments. Weir (1995) also outlined essential characteristics of well-designed exams, including appropriate difficulty levels, effective discrimination between student performances, and clear and unbiased items—all of which were often lacking in the analyzed exams.

Furthermore, the study revealed significant variations in the alignment of exam items with MLC across subjects. Mathematics exams demonstrated a statistically significant positive residual (R = 3.412), indicating a stronger alignment with MLC standards. In contrast, Aesthetics exams showed a negative residual (R = -2.8), reflecting a weaker alignment. This pattern aligns with findings from Abrham et al. (2011) and Workineh et al. (2014), who noted that primary school teachers in the Amhara Region, particularly in subjects like Language and Aesthetics, often lack sufficient competencies in assessing student learning. These findings highlight a critical gap, emphasizing the need for targeted professional development

in assessment practices, especially in subjects traditionally viewed as less technical. Addressing these disparities is essential to ensure consistent alignment with MLC standards across all subjects.

The inadequate performance of Aesthetics teachers in assessment could be attributed to several factors, including the insufficient preparation and training of teacher educators in this field. As noted by Lenburg (2007), the ability of teachers to assess student learning is often shaped by the competence of their teacher educators. In the Amhara Region, many teacher educators in subjects like Music, Art, and Health and Physical Education (HPE) hold only diplomas and may lack the necessary expertise to provide high-quality training in assessment. This suggests a downstream impact from teacher training institutions to primary schools, where teachers may lack the skills to design valid assessments.

One of the more concerning findings of this research was the disconnect between teachers' reported practices and their actual understanding of the MLC. While 100% of the teachers claimed to consider the MLC when developing exams, 6% of the respondents admitted to having no clear understanding of the MLC. This contradiction suggests that teachers may lack a deep understanding of what the MLC entails and how to effectively implement it in exam design. This issue mirrors the challenges identified by Johnston (2015), who pointed to common problems in exam development, such as overly simplistic questions that measure only factual recall, ambiguous wording, and a failure to align exams with course goals and competencies.

The analysis of qualitative data from teachers' open-ended responses revealed four major challenges in exam setting: First, teachers demonstrated a lack of knowledge and understanding regarding how to design exams that align with the Minimum Learning Competencies (MLC). Second, there was a limited availability of material resources needed to create high-quality exams, hindering effective assessment development. Third, many teachers exhibited an unwillingness or lack of readiness to invest the necessary time and effort in preparing exams, which could affect the quality of the assessments. Finally, teachers faced significant administrative challenges, such as insufficient support or guidance in the exam preparation process, which further complicated the creation of well-structured and valid exams.

These challenges are consistent with findings from Oluoch (2014), who highlighted issues such as poorly structured questions, excessive overlap of content, and ambiguous instructions as major obstacles to effective assessment. Similarly, Usman (2016) emphasized the importance of school leadership in ensuring that adequate resources are available for teaching and assessment, pointing to a systemic issue in the broader educational infrastructure that hampers effective assessment practices.

The findings underline the critical need for policy-level support to drive educational reforms in assessment practices. Policymakers should prioritize revising national assessment frameworks to ensure exams comprehensively evaluate students' competencies across all learning domains. Additionally, teacher education institutions must be empowered to equip educators with the skills necessary to design balanced and effective assessments. Furthermore, research into alternative assessment methods, such as project-based learning and performance assessments, could offer valuable insights into more effectively evaluating competencies, particularly in the affective and psychomotor domains. This study lays the groundwork for broader investigations into how educational systems can evolve to deliver more holistic, competency-based assessments that better address the diverse needs of students.

In conclusion, this study underscores the pressing need for comprehensive reforms in both teacher training and assessment practices in Ethiopian primary schools. Teachers must receive clearer guidance on the MLC, along with professional development in designing assessments that address all three learning domains—cognitive, affective, and psychomotor. Furthermore, addressing material and administrative challenges will be crucial in supporting teachers to create high-quality exams that accurately reflect student learning outcomes. These efforts will contribute to the improvement of the overall quality of education and help ensure that assessments truly measure the full range of competencies that students are expected to develop.

5. CONCLUSION AND RECOMMENDATION

This research assessed the quality of Grade Four teacher-made written final exams in seven primary schools in Ethiopia, focusing on their alignment with Minimum Learning Competency (MLC) across five

subjects. Through content analysis, a questionnaire survey, and insights into teachers' exam-setting practices, the study provided a comprehensive understanding of how these exams meet MLC standards and the challenges teachers face in developing assessments.

The findings revealed that over 62% of the exam items failed to meet MLC standards, with a predominant focus on the cognitive domain and a lack of representation of the affective and psychomotor domains. This narrow emphasis on cognitive skills undermines a holistic approach to student development. Notably, while the Mathematics exam better aligned with MLC, other subjects, such as Aesthetics, showed significant deficiencies, illustrating inconsistency in the application of MLC across disciplines. Additionally, the study uncovered gaps in teachers' understanding of MLC and competency- based assessment (CBA) frameworks, exacerbated by limited resources, administrative constraints, and teacher-related challenges, further contributing to misalignment between assessments and learning outcomes.

This research contributes to the broader academic discourse by challenging the assumption that final exams alone can adequately measure MLC. It highlights the need for more comprehensive, contextually relevant assessment methods that capture a full range of competencies. In line with existing literature, the study underscores the importance of designing assessments that go beyond cognitive skills, aligning more closely with local educational contexts and goals. Furthermore, it advocates for a shift in assessment practices that integrates all three learning domains—cognitive, affective, and psychomotor—into a balanced and holistic approach.

The findings have important implications for practice and policy. Policymakers should consider revising national assessment guidelines to ensure that exams more comprehensively measure MLC across subjects and learning domains. Teachers should receive targeted training on competency-based assessment practices, including how to design assessments that reflect the full spectrum of student learning. Additionally, support should be provided for teachers to address the practical challenges of exam development, ensuring they have the resources and knowledge to create balanced, effective assessments.

While this study offers valuable insights, it is not without limitations. The survey was confined to seven primary schools, focused exclusively on Grade Four, and relied primarily on written exams as the sole assessment tool, which may not fully capture the diverse competencies of students. Future research involving a larger, more diverse sample of schools and incorporating alternative assessment methods, such as project-based tasks or practical assessments, would provide a more comprehensive understanding of how to better align assessments with MLC. Additionally, further studies should investigate the impact of teacher training on improving exam design and the effectiveness of revised assessments across varied educational contexts.

Ultimately, this study calls for a structured intervention to improve assessment practices in Ethiopian primary schools and similar contexts. By aligning assessments more closely with comprehensive learning outcomes, the education system can foster more inclusive and effective learning, better preparing students for the challenges of the future.

REFERENCES

Abrham, K., Adisu, L., & Ayehunie, E. (2011). Evaluative research on the professional competencies of diploma graduates of Debre Markos College of Teachers' Education. A research report submitted to and presented at the *Annual Research Symposium of Debre Markos College of Teachers' Education.*

Alhashem, F., & Agha, N. (2020). *Analysis based on the three objective educational domains for final summative secondary examinations of science subjects* (Chemistry, Physics, and Biology).

Anees, S. (2017). *Analysis of assessment levels of students' learning according to the cognitive domain of Bloom's taxonomy.* Institute of Agricultural Extension and Rural Development, Faculty of Social Sciences, University of Agriculture, Faisalabad, Pakistan.

Asnakew, T., & Lake, B. (2015). Teachers' competence in the educational assessment of students: The case of secondary school teachers in the Amhara National Regional State. *The Ethiopian Journal of Education, 35(2)*, 163-191.

Bassellier, G., Reich, B. H., & Benbasat, I. (2001). Information technology competence of business managers: A definition and research model. *Journal of Management Information Systems, 17(4)*, 159-182. https://doi.org/10.2753/MIS0742-1222170407

Baughman, J. A., Brumm, T. J., & Mickelson, S. K. (2012). Student professional development: Competency-based learning and assessment. *The Journal of Technology Studies, 38(2)*, 115-127. https://doi.org/10.21061/jots.v38i2.a.6

Bloom, B. S. (1956). *The cognitive domain: Taxonomy of educational objectives handbook I.* New York: David McKay Co. Inc.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9(2)*, 27-40. https://doi.org/10.3316/QRJ0902027

Carraccio, C., Wolfsthal, S. D., Englander, R., Ferentz, K., & Martin, C. (2002). Shifting paradigms: From Flexner to competencies. *Academic Medicine, 77(5)*, 361-367. https://doi.org/10.1097/00001888-200205000-00003

Carter, M. P., & Williampson, D. (1996). *Questionnaire design.* Staffordshire University, Leek Road, UK.

Cecil, A., & Krohn, B. (2012). The process of developing a competency-based academic curriculum in tourism management. *Journal of Teaching in Travel & Tourism, 12(2)*, 129-145. https://doi.org/10.1080/15313220.2012.674212

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Dave, R. H. (1970). Psychomotor levels. In R. J. Armstrong (Ed.), *Developing and writing behavioral objectives* (pp. 47-60). Educational Educators Press.

Devi, N. B. (2009). Understanding the qualitative and quantitative methods in the context of content analysis. International Conference, Chania Crete, Greece, 26-29 May 2009.

Drisko, J. W. (2014). Competencies and their assessment. *Journal of Social Work Education, 50(3)*, 414-426. https://doi.org/10.1080/10437797.2014.917927

Flanagan, J. C. (1954). *The critical incident technique. Psychological Bulletin, 51(4)*, 327-358. https://doi.org/10.1037/h0061470

Fleishman, E. A. (1953). The description of supervisory behavior. *Journal of Applied Psychology, 37*(1), 1-6. https://doi.org/10.1037/h0056314

Golubina, E., & Loser, A. (2018). *Competency-based exams in professional education.*

Hager, P., Gonczi, A., & Athanasou, J. (2006). General issues about assessment of competence. *Assessment and Evaluation in Higher Education, 31*(3), 303-318. https://doi.org/10.1080/02602930500353161

Hoffmann, T. (1999). The meanings of competency. *Journal of European Industrial Training, 23*(6), 275-285. https://doi.org/10.1108/03090599910284650

Hoge, M. A., Tondora, J., & Marrelli, A. F. (2005). The fundamentals of workforce competency: Implications for behavioral health. *Administration and Policy in Mental Health and Mental Health Services Research, 32*(5/6), 509-531. https://doi.org/10.1007/s10488-005-3264-0

Horton, S. (2000). Introduction: The competency movement: Its origins and impact on the public sector. *The International Journal of Public Sector Management, 13*(4), 306-318. https://doi.org/10.1108/09513550010350331

Hughes, A. (1996). *Testing for language teachers.* Cambridge University Press.

Idris, F., Hassan, Z., Ya'acob, A., Gill, S. K., & MohdAwal, N. A. (2012). The role of education in shaping youth's national identity. *Procedia - Social and Behavioral Sciences, 59*, 443–450. https://doi.org/10.1016/j.sbspro.2012.09.299

Johnston, T. (2015). *Test writing strategies: Opportunities to improve our educational approach.* The Challenges to Writing a Good Test.

Keyton, J., King, T., Mabachi, N. M., Manning, J., Leonard, L. L., & Schill, D. (2004). *Content analysis procedure book.* Lawrence, KS: University of Kansas.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

Lenburg, C. B. (2007). The framework, concepts and methods of the competency outcomes and performance assessment (COPA) model. *Online Journal of Issues in Nursing, 4*(2), 1-12. https://ojin.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume42004/No2Sept04/CompetencyOutcomesandPerformanceAssessment.aspx

McCain, M. (2002). *National skill standards board: Integrating manufacturing skill standards into education and training programs.*

Meyer, T. (1996). *Creating competitiveness through competencies: Currency for the 21st century.* Knowledge Resources Limited.

Ministry of Education (MoE) of Ethiopia. (2003). *Teacher education system overhaul handbook.* Addis Ababa, Ethiopia.

Ministry of Education (MoE) of Ethiopia. (2010). *Curriculum framework for Ethiopian education (KG – Grade 12).* A.M Printing.

Munoz, D. R., & Araya, D. H. (2017). The challenges of competence-based assessment in the educational field. *Educ. Pesqui, Sao Paulo, 43*(4), 1073-1086. http://dx.doi.org/10.1590/S1678-4634201706164230

Nessipbayeva, O. (2012). The competencies of the modern teacher. *ERIC Clearinghouse Publishing*, 148-154.

Neumann, W. (1979). Educational responses to the concern for proficiency. In G. Grant, P. Elbow, T. Ewens, Z. Gamson, W. Kohli, W. Neumann, V. Olesen, & D. Riesman (Eds.), *On competence: A critical analysis of competence-based reforms in higher education* (pp. 66-94). Jossey-Bass.

National Postsecondary Education Cooperative /NPEC/ (2017). *Defining and assessing learning: Exploring competency-based initiatives.* Report of the National Postsecondary Education Cooperative Working Group on Competency-Based Initiatives in Postsecondary Education, Department of Education, USA.

Oluoch, E. (2014). Challenges faced by tutors in setting examinations. *Journal of Education and Practice, 5*(17), 60-73. https://www.iiste.org/Journals/index.php/JEP/article/view/13785

Ozturk, I. (2001). The role of education in economic development: A theoretical perspective. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.1137541

Parveen, H., & Showkat, N. (2017). *Content analysis.* https://www.researchgate.net/publication/318815342

Roegiers, X. (2016). *A conceptual framework for competencies' assessment.* UNESCO IBE (International Bureau of Education).

Rovai, A. P., Wighting, M. J., Baker, J. D., & Grooms, L. D. (2009). Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings. *The Internet and Higher Education, 12*(1), 7-13. https://doi.org/10.1016/j.iheduc.2008.10.002

Roy, G. (2016). *Competency-based assessment in primary education in Bangladesh: A review.*

Schultz, D. P., & Schultz, S. E. (1990). *Psychology and industry today* (5th ed.). Macmillan Publishing Company.

Shannon, D., & Bradshaw, C. (2002). A comparison of response rate, response time, and costs of mail and electronic surveys. *The Journal of Experimental Education, 70*(2), 179-187. https://doi.org/10.1080/00220970209599505

Shippmann, J. S., Ash, R. A., Battista, M., Carr, L., Eyde, L. D., & Hesketh, B. (2000). The practice of competency modeling. *Personnel Psychology, 53*(3), 703-740. https://doi.org/10.1111/j.1744-6570.2000.tb00220.x

Simpson, E. J. (1972). *The classification of educational objectives in the psychomotor domain.* Gryphon House.

Singapore Workforce Development Agency (SWDA). (2012). *Quality assurance division: Develop competency-based assessment plans version.*

Taylor, F. W. (1911). *The principles of scientific management.* Harper & Brothers.

Thomas, A. A., & Cross, K. P. (2010). *Classroom assessment techniques: Taken from the classroom assessment techniques: A handbook for college teachers* (2nd ed.).

Tuxworth, E. (1989). Competence-based education and training: Background and origins. In J. W. Burke (Ed.), *Competence-based education and training: Background and origins* (pp. 1-22). The Falmer Press.

UNESCO. (2001). *UNESCO - Dhaka Ahsania Mission: Training manual on competency-based learning assessment.* Bangladesh.

Usman, Y. D. (2016). Educational resources: An integral component for effective school administration in Nigeria. *Research on Humanities and Social Sciences, 6*(13), 27-37. https://www.iiste.org/Journals/index.php/RHSS/article/view/31294

Weir, C. (1995). *Understanding and developing language tests.* Phoenix ELT Press.

Workineh, K., Dilnesaw, B., Eyasu, A., Mebratu, Z., & Ayrikim, A. (2014). Investigating the professional competencies of teachers graduated with a diploma from Injibara College of Teachers' Education from 2011-2013, Amhara National Regional State, Ethiopia. Research presented at the *Injibara College of Teachers' Education 3rd Annual Educational Research Symposium*, March 4-6, 2015 (Unpublished).