



Assessing Educational Standards and Psychometric Properties of the Certification Test for Primary School Teachers and its Implications for Teacher Performance. The case of Shashemenne City Administration, Ethiopia

KITE BADASA CHURUKE

Department of Psychology. Hawassa University

Hawassa University, Ethiopia

Abstract

The main purpose of this study was to assess educational standards, alignments and educational implications of certification test for primary school teachers in Oromia regional state. A mixed research design was employed, and data were collected from primary school teachers, education office experts and the subject teachers from college of teachers' education in the 2024 academic year. The participants were selected using multistage sampling techniques. The research was carried out based on analysis of primary data obtained from examination (N=25) and interview guide and secondary data obtained from reviewing documents (N= 5). There is a moderate inter-rater reliability for academic content standards ($K = .5985$) and fair agreement between the raters for the alignment of subject objectives and exam items ($k = .219$) and items importance evaluation ($k = .2482$). Thus, in all alignment evaluations, there are supportive values of agreement obtained between the raters that indicate the evidence for the content validity of the certification exam. It also indicated, there are significant value of internal consistence of each domain and entire items of the exam, $\alpha = .808$ for content, $.813$ for pedagogy and $.876$ for total exam respectively. However, there is insignificant and low association between test core and teaching performance. As the results from the interview panel, the implementation of licensing system for the in-service teachers is not effective. Therefore, it is important to internalize its goals within teachers, building positive perception on teachers, and to incorporate all elements of education system and all subjects of the level in the studies.

Key words: educational standards, alignment, licensure test for teachers and educational implication.

Introduction

Education is a tool to produce capable citizens and bring all rounded development of the nation. It plays a great role in contributing to the economic development of both developed and developing countries. Literature reveals that education has a vital role to play in achieving the social and economic progress in every country (Samuel Baah-Duodu, 2023). In Ethiopia, education is conceived as a life to be lived and a means of preparing young people for future life (MoE, 2021:28). Buabeng, Danso & Deodat. (2020) stated that the educational system of any nation serves as a mirror through which the image of its future can be seen and shaped because it is from the school systems that the workforce of the nation are obtained. Therefore, nation's quest at becoming an industrialized country depends on the quality of its citizens who are a product of its educational system. An adequate educational system would always have a capability to grow from time to time based on actual need and it has a parallel process with assessment program. In this regard, assessment is expected to be a bridge in educational system.

Assessments reflect standard and basic competencies which would be obtained by the students. Tests enable policy makers to use them as effective tools for controlling education system. Petric 1987:175 cited in Liying (2000) stated that evaluation and testing has become the engine for implementing educational policy. In educational system, the tests are tools of educational improvement and accountability, and also used to encourage teaching and learning process for quality education. Thus, the educational assessment should reveal the educational standards in terms of its content alignments and performance measure that considered as the attainment of the expected educational outcomes. In this regard, educational standard of the assessment is the extents to which the assessments can measures and align with the components of education system. Gunilla (2008) stated that standard as a components of education system and descriptions in policy documents, defining what the students are expected to know and be able to do as well as how well the students are expected to attain this knowledge and skills. In his review, Gunilla noticed three meanings of standards as common in education: 1) standards as quality indicators; 2) standards as descriptions; and 3) standards in terms of performance standards. Standards as quality indicators deal with control of resources, personnel and business in order to give all students an opportunity to attain the expectations and also deal with the productivity of schools and the whole educational system. A standard as descriptions of what students are expected to know and be able to do, is the most recent meaning of standards in education while the third meaning of standards is connected to standard-setting and is most often called performance standards. Performance standards are concerned with both how well students should attain specific knowledge and skills for different performance levels and a cut-score for the respective performance level on a specific assessment. Thus, the assessment serves as instrument to measure such education standards.

A test which aligns with the national curriculum and the policy directives is essential for quality education. Besides, teacher certification test is used as a means of assuring capable teachers for maintaining quality education

based on the national teacher professional standards. Standards for teachers performance assessment helps to maintain the quality of teaching in respective curriculum and context. In the process of assuring quality education, standard should be taken as the norms, expectations and specifications adopted (Harvey and Newton; 2004).

Literature indicates teacher quality and availability are among the determinants of quality education. Teachers are the executors and facilitators of education. Improving the qualification and competence of teachers is very important for improving the quality of education (MoE, 2021). Satisfactory educational management with high quality in the country improvement up to the wealth, economical achievement, developed society, stable politic, well-known culture, and technological improvement, significantly had to rely upon teachers (Kim and Methee, 2019). Quality teachers are those who exhibit desirable traits that bring about student learning. As mentioned in Adeyemo (2021), teachers' quality includes teachers' teaching experience, the extent of their preparation in subject matter, qualification in area of expertise and their ongoing professional development. It also expected that primary school teachers are required to be equipped with the skills of helping children, tolerance in dealing with children, good academic competence and possess behavior that can make them be models for children (UNESCO, 2013). Teaching force access and quality considered as a key pillar of quality education. It recognized that attention first needs to be given to the teaching force, to issues of training, deployment, continuous professional development (CPD), and working and living conditions, along with teacher accountability to quality teaching (AUC, 2016).

Policy Direction on Teachers Quality and assessment

Ethiopia's aspiring proposal for high-quality teacher education was first set in Education and Training Policy which stipulated that teachers should be certified before being assigned to teach at any level of education and that teachers of all levels of education are required to have the necessary teaching qualification and competency in all repertoire of teaching (MoE, 1994). To achieve the intended goal, education policy has responsibility to maintain teacher education and training for the delivery of quality education. The concern of teacher training policy and the training curricula are to equip teachers with knowledge and skills that necessarily they practice in the classroom. In the Ethiopian education policy, it indicated that teacher trainees should have the ability, diligence, professional interest, and physical and mental fitness appropriate for the profession (MoE, 1994). Thus, the policy statement and its implementation need special attention. There is an effort taken to consolidate quality education based on the policy direction. The existing education and training policy of Ethiopia and education sector documents support the strategies and delivery of quality education. Additionally, a series of ESDPs was designed and implemented from 1997/8 and continued to be updated for the subsequent 25 years that the subsequent documents were revised based on the national economic, social and political desire.

Furthermore, the Ethiopian Ministry of Education introduced the new Education Roadmap in 2018 as a policy of education and training to revise a curriculum and restructuring the education system. In these documents the issue of quality education, relevance, and equity are given due attention. In this regard, educational policies are

likely to be more successful when much effort is done to improve the quality and educational attainment. One aspect of the effort is designing relevant program, teacher preparation and certification criteria for making decisions about individual as well as program as a general.

As the priority issues of education strategies, the TDP2 also given emphasize to teachers professional development in both pre-service and in-service training. Moreover, the current Education Sector Development Plan, ESDP-VI, (MOE, 2021) was designed to strongly address the gaps of the previous EDSPs. Giving emphasis on quality improvement and relevance to the labor market and preparing students for the labor market through a career and technical education (Chala and Shashi, 2022). Alongside with the quality education on base of promoting citizenship behavior among the students, the program committed towards providing quality teaching and teachers by the end of 2025.

The recent Ethiopian Education Development Roadmap document clearly indicates the direction for teacher education:

Establish a comprehensive quality assurance system for teacher education institutes and teacher education graduates; Introduce nationwide quality indicators for placement and exit tests administered in teacher education; Promote an authentic assessment scheme during teacher preparation; and introduce a pass mark policy to be used across the nation; Improve quality of assessment tools or strategies (test item quality, test item analysis, alignment with outcome, etc.); Licensing examinations must be prepared by professionals in the area (in collaboration with experts in educational testing and measurement); Improve teachers' performance evaluation practices; and use performance in licensing examinations as inputs for career development of teachers; and Harmonize teacher deployment and transfer (MoE, 2018:42- 46).

Throughout all the above mentioned education development plans, there is a need to maintain and strengthen national learning and professional assessments for improvements. Besides, teachers' preparation, development and certification have been marked as one of the strategies of maintaining and enhancing education quality. That is way efforts paid to maintain quality of teaching practice. As stated in Samuel Baah-Duodu (2023), teacher licensing is an important element to confirm quality in the teaching world. Licensing establishes the premise to distinguish between prospective teachers with desired level of competence to begin practice and disallowing those who do not qualify from entering the classroom. Likewise, Phelps and Sykes (2020) stated that licensure exams function as a means of supporting claims to distinct expertise and capability, which leads to public trust and recognition of professional status.

The use of educational standards, norms, and determining factors for an educational success reflects the government concern for quality education. Education policymakers are increasingly employing teacher licensing test as a vehicle to ensure a capable citizen that a public- school teachers possess requisite knowledge and skills. Such tests are now being used as licensing examination that is, examination which must be passed by teacher education graduates and working teachers in order to be certificated. The teachers have to be assessed as meeting the national teaching standards and in order to secure their license to teach. Literature indicates the licensing and re-licensing system is a system to improve the standard of teachers and provide incentives for higher skills and outcomes. It sets the national criteria for confirming the capacity of teachers at their development levels. If teachers pass the each level's criteria which include subject knowledge and pedagogy in every 2 to 5 years, they receive re-licenses from the MOE (JICA, 2012). In practice, all regional states have similar certification requirements based on the national teacher qualification frameworks and they set their certification requirements. They require teachers both to complete an approved preparation program, teaching experience (performance assessment) and to pass a certification exam.

As the requirements of the quality education, teachers licensing and qualification system has been given due attention in the policy documents. Since quality of the teaching force is one of the key factors in promoting quality education, only those who possess the required qualities and best performing candidates in national and entrance examinations should be considered for admission to the profession (MoE, 2017: 18; Samuel Baah-Duodu, 2023). For the in-service program the licensure exam also given in the same manner and helps in recognizing teachers professional practices. It is believed that both licensing systems are help to assure the expected teacher competences and develop professional skills. According to the World Bank (2020:52) report document, the teacher licensing system aimed to regulate the professional development of teachers and school leaders; to promote excellence in the teaching profession and in school leadership and management; to identify capacity gaps of teachers and school leaders and provide feedback.

The responsible organ was established under the MoE and manages the system with the collaboration of regional education offices. Consecutively, each region has to put in place quality assurance systems and procedures that meet the needs and the culture of the local society. They regulate the teaching profession through teacher certification programs that serve as gateways into the teaching profession since 2006. Every region has its own procedures for certifying teachers, and every public school is expected to employee teachers that certified by the regional state based on the national teacher professional standards. In Ethiopia, teachers are also required to renew licenses as illegibility for career progression and hence for incentives and salary increment. For in-service teacher professional development there is a policy decision to link career progression with the licensure system as discussed above. The licensing and relicensing systems are mainly based on the test result and work performance in the teaching practice.

This certification test is used as an evaluative type of educational assessment to determine the competence of the teachers. It consists of two major components; i.e. subject area and pedagogical knowledge of the teachers that used to certify them. The assumption is giving the certification test for teachers assumed to be a good indicator for competitive teachers in the work area and till it is serving as a screening device for newly employed teachers and licensing their professional practice to achieve standards set at national level for quality education. Educational experts at the regional level believe that achieving a passing score on specified teacher certification test assumed as placing a qualified, competent teacher in the classroom. A great weight of decisions is being made based on this assumption.

Education standards of the test

Literature indicates educational standard for assessment is an expectation of achievements that maintained within a qualification or subjects. As Gunilla (2008) stated standard refers the content standards that the students are expected to know and are able to do. This component of standards for each corresponding subject and grade level has to be aligned with national assessments. According to AERA, APA, and NCME (1999) the alignment between an assessment and a set of content standards in a subject area has long been recognized as evidence of the assessment's validity-the degree to which the interpretations of an assessment's results can be considered accurate. Sarah, Camila & Peggy (2018) also emphasized on the common expectation and alignment with instruction content and learning outcome. As to these authors, a common set of standards will enable a system of education to better assess, understand, and communicate students' content knowledge and skills. The expectation is that policy-makers will be able to make data-informed decisions to improve academic learning when instruction is aligned with standards and when learning outcomes are assessed on aligned tests.

In the context of education, researchers perceive that the process of establishing the relationship between a test and content standards is required for any assessment to be considered rigorous, high quality, and valid (Case, Jorgensen, and Zucker , 2004). Furthermore, Firdissa (2023, citing Cunningham, Key, and Capron, 2016) have the view that aligning the learning activities to the competencies, objectives, and assessments is essential. Along this, clearly defined competencies need to be linked to assessments, and assessments need to be linked to external measures like career-relevant learning outcomes, which some sources consider as curriculum alignment. According to the report of Ethiopian MoE (2017) for students to attain the expected level of learning effectively, it is necessary that the three components of the teaching/learning process, that is, curriculum, classroom teaching and learning assessment, should be consistent in their principles and contents. Only when these three components share the same set of principles and contents, the teaching/learning process as a whole can be effective and high-quality education can be achieved (Firdissa, 2023). Therefore, from the above discussion and the education perspectives, the educational standards of the certification test can be determined in terms of content alignments and expected performance level. However, there is no comprehensive study that shows the linkage between subject contents and test items prepared for certification.

The Ethiopian Education Development Roadmap (MOE, 2018) indicates the policy practice gap and lack of alignment across policy instruments that the government issues to schools are among others lists of education quality factors. In general, the document sees it as a problem of knowledge transfer in the process of implementation of a state designed curriculum. There is a gap between the demand and skilled human power. Literature indicates, in Ethiopia, there is a significant gap still exists between skills, knowledge and attitudes are relevant to preparing the graduates for careers as professional labour and due care were the most important for make them skilled (Chala and Kant, 2022; Dawit ,2023)).

The USAID (2020) report on mother tongue teachers' competence assessment shows that, in some region teachers are not well qualified to act as mother tongue teachers and obtained diploma without going through proper training. Teachers are hired without assuring the policy requirements. Moreover, for better prediction of success and screening/promoting of the respective teachers who are academically competent, establishing empirically valid and reliable selection device is of a great importance. Taking such steps may help in decision making concerning the employing or promoting of respective teachers. But there is no thorough investigation conducted in the study area. Concerning the implementation, literature shows stakeholder hold different views on certification test. Study conducted in Ethiopia by Olkeba and his colleagues (2019) on the analysis of the teacher training system with specific reference to areas for improvement indicated that the stakeholders were uncertain of qualification requirements for teacher development and certification test.

There are some teachers that failed certification exam. As reviewed in Dawit (2023) in licensing and re-licensing of teacher education college graduates, teachers, and school leaders showed that only 24% managed to pass paper and pencil exams administered after graduation, indicating that most of the graduates from teacher education colleges do not have the necessary academic competence to teach in schools (MoE, 2021). Though licensure exam may not be ensure the minimum knowledge, skills and attitudes necessary to deliver effectively in primary schools, the appropriateness and technical quality of test items need to be ensured as alternative for developing and assessing teacher's' competence. However, there is no study conducted on the effectiveness and item appropriateness of licensure test and the relationship between test score and teachers' teaching performance evaluation.

This is the main gap of research that has to be investigated. Thus, this study aimed to evaluate the educational standards, psychometric characteristic and educational implication of certification test for primary school teachers. Specifically, the study intended to answer different basic research questions. Is the preparation of teacher certification test for primary school teachers in line with the content standards and subject area objectives? To what extent do hiring authorities use the certification test result as promotion criterion? What are the challenges of implementing licensure test for primary school teachers? are among the others that treated in this part.

Research Method and Materials

Research Method

The study followed a mixed method primarily convergent parallel design (Dawadi, Shrestha, & Giri, 2021). In a mixed design, two different data sets namely qualitative and quantitative data are mixed to obtain the triangulated results. It is an approach in which two data sets are combined to get a complete picture of the issue being explored and to validate one set of findings with the other (Creswell and Clark, 2018). Both data sets are complementary, and therefore, there would be an added value in combining them. It uses for one method to elaborate, and clarify the results from another. Besides, in this research, quantitative and qualitative data were collected. First the survey of teacher educator (test administration) and document review were conducted. Then the interview with office experts and college teachers were followed.

The quantitative data obtained from the participants is designed to assess the reliability of the test used for certifying primary school teachers. The data was collected from teachers of primary school in Oromia regional State Shashemene town of the 2024/2016 academic year. The document analysis data used to explore the alignment of test contents and educational standards that determined by respective experts and teachers. The descriptive data obtained from interview was designed to assess the educational implication of the certification test. The researcher also used the interviews to elaborate on the practices, challenges and uses of the certification test result. Data are analyzed separately and the results were interpreted and combine together for drawing conclusion.

Description of the Study Area

Shashemene is a city in southern Ethiopia located in the Oromia regional state. The city lies on the Trans-African Highway from Cairo-Cape Town, about 240km south of Addis Ababa and about 25km north of Hawassa city. As a 2007 national census reported a total population of 100,454, of whom 50,654 were men and 49,800 were women but with an estimated 208,368 inhabitants in the mid 2022. The city has had a municipal status since 1935 and has had a master plan since 1996 <http://www.bath.ac.uk>. Currently it is the administrative capital of the west Arsi zone and is characterized as a level- one city in the region. It is one of the fast urbanizing centers in Oromia Regional State and has organized in to four (4) sub-cities and twelve woredas according to the present Oromia urban reforms. Namely Melkaa sirba, Allache, Harufa and Bishan gurracha Sub-cities.

Population of the Study

The target populations for the study are teachers working in government primary schools during 2024 at Shashemene city administration. In Shashemene town, there are 59 government and 78 private primary schools. The total numbers of teachers employed in government primary schools are 1180 of whom 568 are males and 612 are females. From the city administration, in order to make it manageable the researcher has selected two sub-city /Allache and Melka sirba and six schools were selected from them using simple random sampling method. In the six selected schools, the number of population (mathematics teachers) and their categories in terms of sex were

identified and using purposive sampling methods the participants were taken. Similarly, experts from zonal and city administration education bureau and instructors from the college were selected purposively for interview purpose. The participants were selected by the researcher based on their consent because the data obtained from them and their documents are the best source of information for this study. The sample for this study includes 25 in-service teachers in the diploma level of the teaching qualification. 2 respective instructors from the stream and 3 office experts were taken as participants.

Sampling Procedures

There is a multi-stage sampling procedure. First, the list of sub-city and woredas and sampling frame of the population were obtained from the city administration education office and then the numbers are arranged in teaching field groups of the population. From the list, two sub-cities and one woreda and 3 schools of the participants from each sub-city were selected using probability sampling techniques. To select participants from each school, purposive sampling was employed. Teachers who were selected to participate in the study provided with the certification test, which also included the brief biographical information sheet. The teacher participants were requested to write their identification numbers to enable the researcher to match their responses for analysis.

Instruments of the Study

The study has involved different types of instruments for collecting quantitative and qualitative data from the participants. Mainly, test administration, interview guide and document analysis were employed. In addition to the biographical information collected in terms of self-report assessment (sex, qualification, and teaching experience) achievement test scores was collected. Thus, the primary instrument in this study was examination (academic test prepared for the certification /licensing purpose).

i. Examination (License Test)

Exam is the main instrument of the study that used to collect quantitative data from the participants. It is a type of mathematics academic achievement test and termed as certification examination prepared by MoE with the collaboration of regional education office in 2024 academic year to measure teachers' knowledge, and skills. It is unstandardized test. The test consists of both subject content (maths knowledge 75%) and pedagogical components (professional 25%). The test was reorganized and administered for respective participant teachers of the city and the results were used for quantitative analysis. It consists of 60 total items and each has the expected correct answer. The items are both objective and short answer types. The test was developed assumed to align with the subject content and teaching profession standards. The administered items were dichotomously scored as (0 = incorrect response, 1 = correct response) with a high total score reflecting a high level of teacher competence. The answer for each item will be obtained from the regional Education office.

ii. Interview Guide

The interview guide is the data collection instrument that constructed by the researcher himself after thoroughly reviewing the theoretical and practical aspects of teacher certification test for primary school teachers. The questions of interview guide were prepared in the form of both close and open ended questions type to extract the required data from the participants. The questions then revised and edited by staff members of Psychology department. The questions were target on the practices, challenges and educational implication of the certification test for the quality of Ethiopian education system. It was designed to collect information from the education office experts and selected instructors from the college of teachers' education. Finally, the appropriateness of the interview questions was evaluated and approved by research advisors.

iii. Document Analysis

It is a review by experts. As the name implies, it relies on the opinions of experts/specialists who are knowledgeable about the content covered by the subject and about the development of certification test (education office experts and college instructors). Generally, this method was used to analyze the alignment between examination and subject content and professional standards. First cycle primary school mathematics subject has four areas of learning: numeracy, measurements, shapes, and data handling (Antene, Mulugeta and Kassa, 2021). There are also seven standards that have to be fulfilled by teachers as a teaching competence including subject Knowledge and pedagogical skills. Thus, based on these content categories and professional standards, the alignment analysis worksheet was prepared for the data collection. Mainly, as the alignment analysis methodologies reviewed in Betsy et.al, 2004 the experts review was conducted. The process consists of the systematic item-by-item review of an assessment by experts and content specialists. The review evaluates the alignment of each item and their importance for measuring teachers' knowledge and skills in the area. Before proceed to completing their ratings, all coders were oriented how to identify the exam items related with the subject content standards and curriculum objectives (i.e., expected performance). This orientation included a review of the four subject content standards for mathematics grade 1-4, pedagogical skills and corresponding objectives. Specific descriptions for each of the subject domains covered by the licensure exam were given emphasize. An additional attachment (examination sheet) was given with the evaluation sheet to assist the coders in assigning values for each item. They read each item and code /rate them according to the criteria given on the evaluation sheet. After assigning the categorical concurrence for content standards of each item, the reviewers coded the sheet by identifying the one most appropriate objective among the list that corresponded to the item to indicate the performance of examinees. Finally, they evaluate each item on the three likert-scales of measures deciding whether the items are essential, useful or not necessary, that coded as 3, 2 and 1 for decision of the raters.

Procedure of Data Collection

This study is both a survey and document analysis study. It deals with the effect of teacher certification test for primary school teachers. The quantitative and qualitative data from the participants were directly collected carefully by the researcher using the instruments. Initially, the information required for this study were obtained from the city education offices which includes list of schools, numbers of participants and their qualified departments (streams). Then, data was collected from teachers, office experts and the college instructors. First information regarding teacher's certification test score and the biographical data were obtained from the teacher participants by administering the test in each respective school. The test was administered with the collaboration of school principals.

Similarly, the document analysis process was facilitated by researcher and the items review was conducted by the experts and content specialists. There were orientation for the experts on the evaluation of alignments for content standards and assessment items. Then the expert reviewer expected to fill and rate each item on the format prepared for this evaluation purpose. In the content assessment, they assign item categories on the worksheet containing tables in which each rows of the table have list of items under each standard, objectives and category. They examine carefully whether each item are aligned with standards and decide whether they are essential or not. The qualitative data was obtained from interview with education office experts and from the subject specialists.

Method of Data Analysis

The analysis of data was conducted by using SPSS version 26. To show the general picture of the data, descriptive statistics (mean, variance and standard deviation) and correlation were computed for all participants.

In the process of assessment, the Pearson product moment correlation coefficient was computed between the components of certification test score to show the association with the subject knowledge and pedagogical skills. For knowledge test item type, Cronbach's Alpha Coefficient or coefficient of reliability was computed to see the internal consistency of the test. The item evaluation data from the expert decision will be analyzed for determining test appropriateness and content validity. The data from interview was analyzed in qualitative way to substantiate the quantitative data in exploring the challenges and extent to which the educational authority use test results for assuring education quality. Finally, based on the results of the study discussion, conclusion and recommendation were drawn.

Ethical Considerations

In researches that have relationship with social, political and economic implications, ethical consideration is obligatory. Throughout the whole process of the research, the researcher respects the ethical principles and rules of his profession. Freedom of responses, reply and withdraw from the research at any time had given priority and informed before any single progress of research process. Besides, the researcher highly respects the culture,

language, religious perceptions, and any form of life style of the participants on which the research was conducted. Also, the researcher promised to keep all information / data collected from respondents confidential.

Result and Discussions

The study was conducted to attest the educational standards and psychometric characteristic of the licensure exam for primary school teachers. The content alignments and reliability of the exam and the importance of test items for the licensing purpose were evaluated. The analysis focused on the coverage of examination with the core learning contents and expected learning outcomes of the mathematic subject for primary schools. It tried to assess the inter-rater reliability of the test items coded by the alignment reviewers.

Table 1. Demographic characteristics of the participants

Groups of the participants	Sex			Academic rank			Experiences			Licensed	
	Male	Female	Total	Dip	1 st deg.	2 nd deg.	1-4	5-9	>10	Yes	No
Alignment evaluation panel	4	1	5	-	2	3	-	2	3		
Teachers participants	22	3	25	18	7	-	2	9	14	6	19
Total	26	4	30	18	9	3	2	11	17	6	19

As indicated in the table 1, majority of the participants are males 26(86.6%) and diploma level of academic rank 18(60%) and majority of them 17 (57%) have more than 10 years of work experiences among the two groups. It also show that among the teacher participants 19 (76%) were not licensed. Only 6 (24%) teacher participants were licensed through the in-service licensure system.

Members of the alignment evaluation panel have only first and second degree academic rank and they had extensive understanding of the subject content, testing and measurement, teacher licensing system, and education policy as general. The majority of panel members also had participated in the teacher licensing examination, and they were familiar with the instrument and its application before the alignment assessment.

Subsequent analyses of the standard alignment ratings resulted in descriptive statistics for the content and objectives alignment model: a categorical concurrence and Range of knowledge alignment model and the reliability of inter-rater agreement for the entire examination were computed. In the analysis the average value coded by all reviewers was considered as the reference for alignment evaluation.

First, as per the required criteria, the average number of items in the exam that the raters coded as corresponding to a content standards and the number of items in each standards rated as essential or acceptable for the purpose it was prepared. Second, the inter-rater reliability for total exam items and test items reliability for each components /domains/ and total exam items are presented as follows.

Table 2. *Categorical concurrence /alignment/ for the academic standards and item importance evaluation*

Subject domain /prof. standard	No of items in licensure exam for the domain	Academic content standards	Average no of items in licensure exam the raters coded as corresponding to an academic standard (%)	A number of items in academic standards rated as essential or acceptable (%)
Mathematics content Knowledge	37	Number system and operations	10 (27%)	9 (90%)
		Measurement	15 (40.5)	12 (80%)
		Shape	6 (16.2%)	4 (66.7%)
		Data handling	6 (16.2%)	5 (83.3%)
Mathematics pedagogical skills	23	Understand how students learning	3 (13.04%)	3 (100%)
		knowledge of math teaching strategies	7 (30.4%)	7 (100%)
		Knowledge of planning	7 (30.4%)	6 (85.71%)
		knowledge of assessment and giving feedback	6 (26.08%)	6 (100%)

As indicated in Table 2, except one pedagogical standards (Understand how students learning), all are coded with 6 and above numbers of items and majority of items in each standards are rated as essential for the purpose they were prepared. Only few numbers are considered as useful and no items that rated as not necessary. For example, there are 10 items 27% of the domain category coded for number system and operations which 9 (90%) of them rated as essential and one as useful whereas, 15 items (40.5% of the domain category) coded for Measurement which 12 (80%) of them rated as essential and 3 (20%) as useful items.

The results also indicate that the inter-rater agreement and scale reliability for the entire items of the exam are acceptable. In the analysis, the possible pair agreements of all coders for each evaluation are computed. Then the average values of all agreement are presented in the table below. Assuming the null hypothesis, the Kappa agreement values of academic standards/category for each pair of coders are moderate ranged from $k = 0.495$ to 0.687 with the significant Pearson Chi-square values $P < .05$. For the subject learning objectives and item evaluation were also computed and significant values obtained for each. The Kappa agreement values ranged from $.055$ to $.334$ and $-.026$ to $.563$ respectively. All kappa values are different from 0, i.e the observed agreement between raters are not caused by chance. Moreover, since the measurement was nominal/ordinal scale and one sample used for testing occurrence, the non-parametric Chi- square test was then performed for entire items of the exam. The findings indicate that the occurrence of mean values with equal probability is not significant $P > 0.05$ for academic content standards and subject objectives, while significant for the items importance evaluation $P < 0.05$. Therefore, these confirm that for academic standards and learning objectives there are probability of equal occurrence of

values for each rater /the null hypotheses is retained and not for the items importance evaluation (i.e there are no probability of equal occurrence of values for each rater on item importance evaluation).

Regarding the test importance and its implementation, the participants have different views and perception towards the test and the whole certification process. As the data from the interview with office experts indicated the implementation of licensing system for the in-service teachers is not effective.

Response for the questions on preparation, content, importance and application of the licensure exam indicates:

The exam was prepared by subject matter specialists and assessment experts coordinated at the national level, MoE. Since the exam expected to cover the contents in the range grade levels (1-4) I don't believe to touch each topics of the grades. But as a category of content standards it tried to incorporate all standards [X3].

Even the aim of the exam is good there are problems with implementation. The use of criteria for licensing (exam and work performance) are not continuously implemented, there is inconsistencies of exam administration and teacher's career progression that makes the system difficulty to implement [X1].

Despite of interruption on implementation, all interviewees respond the some on the importance of license exam. The response of other interview X2 on the challenges of certification process says:

Due to different reasons most teachers have no interest and readiness to take the licensure exam; even they have low work performance due to their personal and contextual influences. For example, economic and social life of the teachers makes them discouraged and daunted from the professional work and they look for other options and desist from academically upgrading and professional development. Most teachers wish to change their work; they have no feeling to persist in the profession. Therefore, it requires more attention from government and all stakeholders.

Other respondent of the office expertise X4 said that the extent to which the authorities used the certification results for teacher promotion and professional development was very low. Because of the interrupted licensing exam administration in the last three years and few numbers of teachers that achieve pass mark on the exam, few numbers of teachers were promoted based on their exam results. The remaining majority teachers were below the expected standards. These indicate that there are contextual and personal challenges for implementing the licensing system for the teachers at primary school level.

Furthermore, the exam and items evaluation result shows that there are acceptable levels of internal consistencies of the exam items.

Table 3. *Inter-rater reliability of the alignment and the non- parametric tests for each evaluation*

Computed items	No of the coders/ raters	Average Inter-rater Reliability (Kappa agreement)	Hypothesis /one sample chi-square tests	
			p-values	Decisions

Academic content standards	5	.5985	.320	Retain the hull hypothesis
Subject learning objectives	5	.2190	.458	Retain the hull hypothesis
Items importance evaluation	5	.2482	.000	Not retain the hull hypothesis

Various interpretations for kappa can be found in the literature, such as those proposed by Landis and Koch: 0.8–1.0 indicating almost perfect agreement, 0.6–0.8 substantial agreement, 0.4–0.6 moderate agreement, 0.2–0.4 fair agreement, zero to 0.2 slight agreement and zero or lower poor agreement (as cited in Joachim Pum, 2019). From the above table 3, it indicated that there is a moderate inter-rater reliability for academic content standards ($K = .5985$) and fair agreement between the raters for the alignment of subject objectives and exam items ($k = .219$) and items importance evaluation ($k = .2482$). Thus, in all alignment evaluation there are reliable and supportive values of agreement obtained between the raters that indicate the evidence for the content validity of the certification exam.

Interview results also indicates even the contents to be covered are large, the content standards are included and respective learning objective are considered [X1]. Similarly, other respondent [X4] said that though the complexity of contexts and large contents across grades, the license exam prepared in line with expected learning outcomes and content standards.

Table 4. Test items reliability ($N = 25$)

Exam domains /prof. standard	No of items for the domain	Reliability	Standard deviation
Mathematics content Knowledge	37	.808	5.394
Mathematics pedagogical skills	23	.813	4.443
Total exam items	60	.876	8.741

The reliability of items for each domain /category was computed on the data obtained from 25 teachers currently teaching mathematics in primary schools. As the result depicted in table 4, the reliability of each domain of the subject and entire items of the exam are at acceptable level alpha = .808 for content knowledge with 5.39 SD, .813 for pedagogical skills with 4.44 SD and .876 for total exam items with 8.74 SD respectively. There are no rejected items and the corrected item- total correlations are medium and only low for items 10, 36, 37, 47 and 52 that the Cronbach alphas slightly surpass the total reliability of the scale, when the items deleted. Generally, results indicated that the total reliability of the licensure exam and the subscales reliabilities are at acceptable level.

The relations between test score, teaching performance and CGPA

As shown in Table 5. the relationship between teachers academic ability measures and teaching performance indicates there is positive, low and insignificant correlation between certification test score and teaching performance $r = .369$, $P > .05$ and low and insignificant relation between certification test score and college grade point $r = .337$ $P > .05$, these may resulted from the context and low preparation of the teachers for the examination. However, a medium significant relation was found between teaching performance and college grade point $r = .429$, $P < .05$.

Table 5. Correlations between test score, teaching performance and CGPA

	Total score on certification test	Teaching performance	College achievement (CGPA)
Total score on certification test	1		
Teaching performance	.369	1	
College achievement (CGPA)	.337	.429*	1

* Correlation is significant at the 0.05 level (2-tailed).

The data reveal that the average score value of certification test is low and blow the expected standards. Mean = 42.08 and SD = 4.93 which is below the minimum requirement as passing score (50%). And relatively high average score in teaching performance score and CGPA, $M = 85.92$ $SD = 4.01$ and $M = 2.9$ $SD = .32$ respectively.

Discussion

This investigation used the theoretical model, Webb's alignment model, which provides a series of interaction that indicate the match between the content in the national academic standards and the content covered by the teachers licensure exam. By focusing on the alignment of the licensure exam with academic subject standards, this study provided content related evidence for the validity of the licensure exam. Specifically, the investigation sought to answer the following question: Is the preparation of teacher certification test for primary teachers in line with the content standards and subject area objectives? It tried to identify whether the certification /licensure exam adequately measure the concepts and skill areas represented in subject content standards. As it indicated in the results the alignment of exam item with the mathematics content standards are medium and statistically significant. However, there are low and significant alignment between exam items and cognitive learning objectives. The coding agreement of the raters implies majority of items are coded as measuring low level cognitive skills as they arranged from simple (knowledge level) to more complex higher level (evaluation/creation). Learning objectives are an expectation of the knowledge, skills and attitude that the learners will achieve and the assessment should prepare in line with them. Therefore the study reveals that the cognitive objectives of the primary mathematics subject are represented in its respective licensure exam prepared for the teachers in 2019 academic year. However,

as the coded data indicates the affective and psychomotor domains are missed and attention was not given to them. From this it's possible to say the certification exam is not fully covered the expected learning outcomes. There are moderate values of interrater agreements on standards, objective and exam. Literature shows learning objectives should align with the assessment. Supporting this finding Firdissa (2023) stated that in principle, teaching-learning should be momentarily guided by clearly articulated learning outcomes; learners' awarded grades should undoubtedly signify learning outcomes; and assessed tasks (knowledge, skills, and attitude) should certainly be aligned to clearly articulate learning outcomes for the levels of achievement. The item and expected outcome alignment is the base for the content standards and determining performance standards.

The small values of agreement may result from the computed numbers of elements and types of the content. As the theoretical assumption the alignment coding process is not designed to produce exact agreement among members of the coders' panel. In fact, variance in ratings may represent valid differences in opinion that reflect a lack of clarity in the objectives or the robustness of assessment items that could reasonably correspond to more than one curricular objective (Webb, 2002 cited in Roach, Elliott and Webb, 2005). Similarly, due to its type and numbers of content standards it may be challenging to construct items that represent all standards. Different types of standards require different formats of assessments and items and some kinds of standards are difficult to assess at all. Such limitations in assessment format may also limit the proportion of assessed standards and thereby limit the highest possible degree of alignment.

It is believed that teacher license exam is designed to know and improve teacher's competence as professional standards for maintaining quality education. The study indicated that despite the overall positive views regarding the teaching licensure, stakeholders failed to recognize its impact on their personal and professional development. Teachers perceived the teaching licensure as difficult, adding to their workload rather than supporting their professional growth. This result is in line with the study conducted by Olkeba and his colleagues, 2019 indicated there is uncertainty on implementation of the certificate exam but it is remarked as important in future teacher training programs.

As the current study indicates, there are low scores of certification test of the participants. This result is consistent with the study conducted by Dawit (2023) and the report of MoE (2021), in licensing and re-licensing of teacher education college graduates, teachers, and school leaders showed that only 24% managed to pass paper and pencil exams majority of them were failed. The study also indicates there is low and not significant relation between certification test score and other teachers' performance assessments. This result is inconsistency with Ghamrawi, Abu-Tinehans and Shal, (2023) reviewed that teacher licensing does have some predictive power in terms of teacher performance in the classroom, that licensed teachers possess the necessary knowledge, skills, and competencies to effectively teach and, thus, contribute to education. The result current study might be due to teacher low preparation for exam, perception, accountability of stakeholders and life conditions of the teachers are among other factors

that make teachers low perform on license exam. Finally, assessing and improving teachers' capability and establishing alignment among curriculum elements is an essential activity for improving the potential for realizing quality education.

Educational Implication

This study can help the policy makers in maintaining the quality of education. As the procedure and results of this study shows the role of sequential development and expert review in promoting the alignment between the elements of education system; curriculum content standard, instructional objectives, and assessment systems in which the educational standards and expected outcomes are serving as base for the development of the assessment items and based on the pre-determined standards and outcomes, reviewing its appropriateness are important. It also indicates the reliable and valid instruments helps to measure the teacher performance and the status of education system as general. Thus, the finding would be applied in showing ways to evaluate the examination and it is very important for the expert review to analyze the importance of each item. This also implies for the total education system in the constructive alignment of all components: the curriculum content, learning objectives, instructions, and assessment strategies. The licensure system serves as the means of regulating education quality and teacher employment system. More importantly, the licensure processes should foster a developmental culture rather than a regulatory one.

Conclusion and Recommendations

Based on the data analysis, the results shows that a considerable effort was made to align content standards, learning objective, and assessment that indicate the effective implementation of teacher certification exam and to maintain quality education. The moderate alignments were found between licenser exam and content standards and subject objectives and weak practice for educational implication; used for teacher promotion. It also found that there is low and insignificant relationship between licensure exam score and teaching performance in the school. However, it has a moderate and significant relation with the college GPA. It is also evident that the overall mean score in the licensure examination for teachers is low and there are contextual, perceptual and accountability challenges identified that need further consideration of concerned bodies. It has, therefore, recommended that: To indicate the total picture of interrelation of education system, it is necessary to incorporate all elements and all curriculum of the level (alignment of contents, learning objective, instruction, and assessment). To achieve the intended goals of teaching licensure exam, it is crucial to internalize its goals within teachers, building positive perception on teachers that enable them to persist and effectively improve their professional practice. Teachers also should prepare themselves to be assessed on what they have taught and on practical performance that they possess as school expectation that help them in improving professional competence and their career development. The Ministry of Education along with the college of teachers' education should put in place vertical and horizontal accountability setups to embed the licensure assessment tasks continuously implemented within the overall

curriculum so that alignment of contents, learning objective, instruction, and assessment can be shared as responsibility of all related stakeholders. Further studies that provide a detail evidence of the psychometric properties of the licensure exam will be necessary to establish the construct validity.

Declaration

This study, entitled "Assessing Educational Standards and Psychometric properties of the Certification Test for Primary School Teachers and its Implications for Teacher Performance," is my original work and has not been published on any educational Journals. I have duly acknowledged all sources of material used in the study.

Acknowledgement: I am grateful to all participants who took their time and energy to provide required information and Hawassa University for financial support.

Conflicts of interest: I declare that there is no conflict of interest.

References

- Adeyemo Emily Oluseyi (2021). Assessing Teachers' Competence in Items Development Through Evidence of Convergent Validity of Test Scores from Alternate Examinations, *International Journal of Psychology and Behavioral Sciences*, Vol. 11 No. 5, 2021, pp. 79- 83. doi: 10.5923/j.ijpbs.20211105.01.
- African Union Commission [AUC]. (2016). Continental Education Strategy for Africa 2016-2025 CESA 16-25. Retrieved from https://au.int/sites/default/files/documents/29958-doc-cesa__english-v9.pdf
- African Union (2019). African Teacher Qualification Framework For Teacher Quality, Comparability and International Mobility. Addis Ababa, Ethiopia
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC:
- Anteneh Tefera, Mulugeta Atnafu and Kassa Michael (2021). The Relevance of Current Ethiopian Primary School Teacher Education Program for Pre-service Mathematics Teacher's Knowledge and Teacher Educator's Awareness about Mathematics Knowledge for Teaching. EURASIA Journal of Mathematics, Science and Technology Education. Research Paper retraived from <https://doi.org/10.29333/ejmste/10858>
- Beth Roseman (2018). From Policy to Practice: Perceptions of Effective Teaching. A Qualitative Study of Public and Private Primary Schools in Addis Ababa. <http://www.duo.uio.no>
- Case, B. Jorgensen, M. and Zucker S. (2004). Alignment in Educational Assessment. Assessment report, Pearson Education, Inc.
- hala Wata Dereso and Shashi Kant (2022). Review of Ethiopia Education Sector Development Programme VI in Perspectives of Filling the Labour Market Skill Gap. Journal of Culture, Society and Development. Vol, 67 An International Peer-reviewed Journal

- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). Sage Publications.
- Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-Methods Research: A Discussion on its Types, Challenges, and Criticisms. *Journal of Practical Studies in Education*, 2(2), 25-36 DOI: <https://doi.org/10.46809/jpse.v2i2.20>
- Federal Democratic Republic Government of Ethiopia (1994). Education and Training Policy St. George Printing Press Addis Ababa
- Firdissa Jebessa (2023). Learning Outcomes, Instruction, and Assessment Alignment: The case at two Ethiopian Universities *International Journal of Research Publications* 135(1), 206-217; doi:10.47119/IJRP10013511020235573Ghamrawi, N.; Abu-Tineh, A.; Shal, T. (2023).Teaching Licensure and Education Quality: Teachers' Perceptions. *Sustainability*, 15, 10886. <https://doi.org/10.3390/su151410886>
- Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, 47(1), 218–251. <https://doi.org/10.3102/0002831209348970>
- Government of Ethiopia (2018a). Ethiopian Education Development Roadmap (2018–30): An Integrated Summary. Ministry of Education, Education Strategy Centre.
- Gunilla Nasstrom (2008). Measurement of Alignment between Standards and Assessment, Unpublished Thesis Printed by Print & Media, Umeå University: 2005262 ISSN 1652-9650
- Harris D. And Sass T. (2006).Teacher Training, Teacher Quality And Student Achievement. Working Paper 3
- Harish, K. et. al, (2016). Measurement and Evaluation in Education. Teaching Material. Tripura University. Vikas publishing House pvt. LtdJames Keevy (2020). African Continental Qualifications Framework ACQF MAPPING STUDY Country Report Working Paper.
- James Cowan, Dan Goldhaber, Zeyu Jin, Roddy Theobald (2020).Teacher Licensure Tests: Barrier or Predictive Tool? CALDER Working Paper No. 245-1020 Washington, DC 20007 202-403-5796. www.caldercenter.org
- JICA (2012). Basic Education Sector Analysis Report. International Development Center of Japan Inc. (IDCJ)
- Kim Dimanchetra and Methee Wisaprom (2019). Factor Analysis of Competencies of Teacher Performance in Secondary Schools in Svayrieng Province, Cambodia. *International Journal of Education and Research* Vol. 7 No. 5
- Maxwell, J. A. (2016). Expanding the history and range of mixed methods research. *Journal of Mixed Methods Research*, 10(1), 12–27. <https://doi.org/10.1177/1558689815571132>Ministry of Education (2015). National Professional Standards for Teachers. National document AddisAbaba.

Ministry of Education (2015). Teacher's Standard Licensing and Relicensing Directorat Professional Standards for Special Needs Education Teachers Grade 1-12 Adama, Ethiopia MoE (2015). Education Sector Development Programme V (ESDP V) 2015/16 – 2019/20 G.C.

Program Action Plan. addis Ababa

Ministry of Education (2017). Ethiopian Education Development Roadmap (2018-30) Education Strategy Center (ESC)

MoE (2017). Project for Capacity Development for Improving Learning Achievement in Mathematics and Science Education in the Federal Democratic Republic of Ethiopia (LAMS) Project Completion Report October 2017 JICA. International Development Center of Japan Inc. Koei research and consulting Inc.

MoE (2021). Education Sector Development Programme VI (ESDP VI) 2020/21 – 2024/25 G.C. Addis Ababa

National Research Council (2001). Testing Teacher Candidates: The Role of Licensure Tests in Improving Teacher Quality. Washington, DC: The National Academies Press. doi: 10.17226/10090. Olkaba, T., Hunde, A., Mamo, T., Duresa, G. and Keno, D. (2019). Analysis of the Teacher Training System in Ethiopia with Specific Reference to Areas for Improvement. Makerere

Journal of Higher Education. V 10 (2), P 157 – 170 DOI:

<http://dx.doi.org/10.4314/majohe.v10i2.12> Phelps G. and Sykes G. (2020). The Practice of Licensure, the

Licensure of Practice KAPPAN

<https://kappanonline.org/practice-teacher-licensure-professionalism-phelps-sykes/> Popham, W. and

Yalow, E. (1984). Standard setting Option for Teacher Competence tests. Paper presented at a Joint Session of the American Educational Research Association and the National council on Measurement in Education. New Orleans. Apr 23-27, 1984 Roach, T. Andrew, Elliott, N. Stephen and Webb, L. Norman (2005). Alignment of an Alternate

Assessment with State Academic Standards: Evidence for the Content Validity of the Wisconsin Alternate Assessment. *The Journal of Special Education* Vol. 38/No. 4 Pp. 218–231

Samuel Baah-Duodu (2023). Teacher Licensing and Valid Acuity to Quality Teaching in Ghana. *Research Highlights in Language, Literature and Education* Vol. 7 doi: 10.9734/bpi/rhll/v7/19468D

Sarah M. Bonner & Camila Torres Rivera & Peggy P. Chen (2017). Standards and Assessment: Coherence from the teacher's perspective. Educ Asse Eval Acc <https://doi.org/10.1007/s11092-017-9272-2>

Webb, N. L. (1997a, April). Research monograph No. 6. Criteria for alignment of expectations and assessments in mathematics and science education. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (1999). Research Monograph No. 18. Alignment of science and mathematics standards and assessments in four states. Washington, DC: Council of Chief State School Officers.

Workneh Abebe and Tassew Woldehanna (2013). Teacher Training and Development in Ethiopia
Improving Education Quality by Developing Teacher Skills, Attitudes and Work Conditions. An International
study of childhood poverty. www.younglives.org.uk

USAID (2020) Mother Tongue Teachers' Competencies Assessment (MTTCA) Reading for Ethiopia's
Achievement Developed, Monitoring and Evaluation (READ M&E) 2019 *Report, Revised June 2020.*

American Institutes for Research NW Washington, DC

20007

