Heliyon 10 (2024) e31869

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Alignment analysis between teacher-made tests with the learning objectives in a selected school of central regional state of Ethiopia

Tsedeke Abate [a,*], Ertembo Mishore [b]

[a] *Department of Physics, Hossana College of Education, P.O. Box 94, Hossana, Ethiopia*
[b] *Department of Curriculum and Instruction, Hossana College of Education, P.O. Box 94, Hossana, Ethiopia*

ARTICLE INFO

Keywords:
Alignment
Analysis
Assessment
Learning objectives
Teacher-made tests
Course objectives
Bloom's taxonomy
Porter's alignment model

ABSTRACT

Alignment studies can offer valuable insights to educators about the effectiveness of the course objectives, assessments, and teaching. In this particular study, the aim was to determine the extent to which teacher-made tests aligned with the learning objectives of natural science subjects. The study included a total of 180 learning objectives (46 from Biology, 71 from Chemistry, and 63 from Physics) and 88 test items (30 from Biology, 30 from Chemistry, and 28 from Physics). Bloom's revised taxonomy was used to identify, organize, and code the objectives and test items. Porter's alignment index was used to analyze the data, allowing for determining the degree of alignment between tests and learning objectives. The results showed that the overall alignment between teacher-made tests with course objectives was 45 % for Biology, 46 % for Chemistry, and 62 % for Physics. The study also found that a dependable degree of alignment was not yet established between learning objectives and tests. Therefore, it was suggested that teachers should use assessment procedures and blueprints that consider higher-order cognitive levels and expected learning objectives. Further investigations are also required to determine whether students are meeting the expected learning objectives and moving to the next grade level.

1. Introduction

According to the Ethiopian Ministry of Education [1], education is crucial for a country to become a knowledge-based society that effectively utilizes new technology to solve present and future problems. However, educational policymakers are currently grappling with the challenge of equipping students with the knowledge, skills, and attitudes necessary to thrive in a globalized 21st century [2]. The National School Curriculum, which outlines a program of learning, is a critical document that reflects the nation's aspirations and prepares citizens for future challenges. Unfortunately, despite initial excitement during its adoption, many have been left disappointed with the implementation of curriculum documents [3].

The Ethiopian Ministry of Education has recently implemented a new curriculum framework aimed at addressing educational challenges in Ethiopia. Among these challenges are inadequate teaching methods and difficulties in assessment implementation, which have contributed to a decline in educational quality and course objectives. As Yambi [4] notes, well-designed assessment strategies play a crucial role in educational decision-making and are essential for ongoing quality improvement processes at the lesson, course, and curriculum levels. Additionally, National Research Council [NRC] [5] emphasizes the importance of aligning educational

* Corresponding author.
E-mail address: tsabge@hec.edu.et (T. Abate).

https://doi.org/10.1016/j.heliyon.2024.e31869
Received 16 November 2023; Received in revised form 10 May 2024; Accepted 21 May 2024
Available online 24 May 2024

assessments with curriculum and instruction to support students' classroom learning.

Alignment studies are a valuable tool for researchers to systematically analyze the various components of an educational system, comparing their content and evaluating their level of agreement [6]. These studies provide insight into the extent to which assessments and standards align with students' classroom learning. Standards refer to descriptions of the knowledge, skills, and abilities that students are expected to develop in a specific subject area [7]. In this study, "standards" refer to the "course objectives" outlined in science curricular materials. Assessments, on the other hand, are tools used to gather data on what students have learned. Examples include summative, formative, and teacher-made tests. For the purposes of this study, "assessment" refers specifically to teacher-made tests, particularly the final exam administered at the end of a semester. These tests are typically designed to measure outcomes and content in line with the national curriculum and are expected to cover the entire content area with a large number of items. Classroom teachers rely on assessments to determine what to teach, how to teach, and what instructional materials are appropriate for students' classroom learning [8]. As such, teacher-made tests play a significant role in maintaining the quality of education. They are used to establish whether students have met the learning objectives, identify areas for improvement, and provide necessary feedback and revisions [9]. Teacher-made tests offer advantages over standardized tests in contributing to the achievement of educational goals, as they are developed directly from the learning objectives of the courses taught in the classroom.

The success of education largely hinges on establishing a significant connection between educational components like learning objectives and the accuracy and effectiveness of corresponding assessments. The Ethiopian curriculum, newly developed, aims to integrate higher-order learning into its course objectives, which include critical thinking, analysis, evaluation, creation, and problem-solving [10]. In light of this, teachers must ensure that there is a match between such higher-order learning and item construction in assessments. Assessments that focus solely on knowledge recall tend to encourage shallow learning, whereas assessments that incorporate low to high levels of learning, from basic to critical thinking and creative problem-solving, tend to lead to higher levels of student performance and achievement [11]. Therefore, teachers must possess the necessary skills and abilities to help their students enhance and develop their higher-order thinking capacities as outlined in curricular documents [12]. It is reasonable to assume that students will attain their expected knowledge, skills, and abilities as stated by national curricular objectives when there is a better alignment between course objectives and teacher-made tests. Conversely, it becomes challenging to gauge a student's learning when assessments and course objectives are not aligned [13].

Currently, in the Ethiopian context, almost all students are getting passing marks to be promoted from one grade to another without any challenge. However, most students are facing a great challenge in scoring pass marks on national exams. For instance, National Learning Assessments (NLA) conducted in 2000/01, 2004/05, and 2007/08 in grades 4 and 8 showed that the composite scores of students are always below the national minimum of 50 % [14]. Besides, the mean scores are going down from 47.9 % to 40.9 % in grade 4 and from 41.1 % to 35.6 % in grade 8 between the first and third NLA. Very poor achievements have also been observed in grades 10 and 12 NLA report of 2010 [15]: only 13.8 % of Grade 10 students and 34.9 % of grade 12 students have passing marks (50 % and above) in the science subjects, Mathematics and English [16]. It is also worth mentioning that the 2022 and 2023 students' poor achievement (only 3.3 % and 3.2 % of students scored a passing mark respectively) in the National grade 12 Examination [17]. As a result, what is happening in the education sector requires immediate attention and adjustment.

To address the issue of low student scores on regional and national exams, one potential solution is to assess how effectively teacher-made tests align with standards, specifically course objectives. It is also important to evaluate tests in terms of Bloom's Taxonomy's cognitive demands, which range from higher-order thinking skills such as analysis, evaluation, and creation to lower-level skills like remembering, understanding, and applying (*For more information on this topic, please refer to the conceptual framework section*). By evaluating whether students are being assessed following the established course objectives, it is possible to gain valuable insights into their performance. However, there is currently a paucity of research at the national levels on the alignment of teacher-made tests with standards. While there is literature supporting the use of a table of specifications or test blueprint to determine the validity of teacher-made tests, there is little information available on how to make these tests compatible with different levels of course objectives [18,19].

To ensure that the newly developed national curriculum is being effectively implemented, it is crucial to gather empirical evidence on the alignment between teacher-made tests and course objectives with the revised course objectives and assessment practices [10]. This is particularly important because a lack of coherence between assessment, instruction, and curriculum has hindered the implementation of a constructivist approach to education, which is the foundation of competency-based learning [20]. Thus, the implementation of a new curriculum framework is imperative to address Ethiopia's ongoing educational challenges.

As has been previously mentioned, a decline in educational quality that has not been adequately noticed over the years has come to light in the public discourse since the 2021/2022 academic year, when the majority of grade 12 students scored a significantly lower achievement on the national examination. In a similar vein, students frequently struggle to pass the regional exams for grade 8. The researchers now have the opportunity to more easily question why students who passed teacher-made examinations from primary school through secondary school did not meet the course objectives set by the regional and national governments. Thus, alignment studies allow researchers to determine whether teaching, assessment, and course objectives are all working together harmoniously to provide reliable teacher assessments of student achievement.

To this end, the researchers' motivation to carry out this study was to find a pertinent solution to this practical problem. Even though the scope and magnitude of this study are small, it is hoped that the results will help one determine the degree of agreement between the two components (teacher-made tests and objectives) and may shed light on the "black box" of classroom instruction, which may reveal some information about what actually happened during teaching and learning in the classroom. This could provide guidance on corrective actions to be taken during the next national curriculum framework implementation, as well as act as an indication for educators to reconsider and modify their assessment procedures in accordance with the stated objectives.

This study aims to determine the degree to which teacher-created tests align with the learning objectives of natural science subjects for 10th-grade students, as well as their cognitive demands. As Townley [21] notes, science education is intended to foster a scientifically literate society, which includes problem-solving, evidence-based, decision-making, and reasoning-based judgment. The researchers, who possess extensive experience in teaching science and education, undertook this investigation to address this issue. To guide the research, the researchers utilized the following questions:

1. To what extent do teacher-made tests align with the learning objectives of science subjects?
2. Are the learning objectives of science subjects in agreement with the six levels of Bloom's cognitive demand?
3. Are the test items constructed by the teachers complement with the six levels of Bloom's cognitive demand?

2. Conceptual framework

In the context of education, alignment could refer to a variety of concepts. It describes the degree of consistency between two or all three of an educational system's constituent parts, namely instruction, assessments, and course objectives [22]. One way to describe this would be that each assessment item corresponds to one or two course objectives [7]. Thus, one way to demonstrate or assess the relationship between curriculum, assessment, and instruction is through alignment research.

Despite the paucity of research on the use of alignment research in the classroom [22], the findings of an alignment study may be able to assist educators, policymakers, and assessment developers in fine-tuning curriculum, assessment, and instruction to better support student expectations. Researchers can systematically examine the various elements of an educational system through alignment studies, comparing their content and evaluating how well they align. Additionally, they offer data on how closely course objectives and assessments correspond with in-class instruction [6].

Course objectives in education refer to the guidelines outlining the skills and knowledge students should know, as well as how well they should demonstrate these abilities [7]. The standards-based reform movement is built on the premise that stakeholders in an educational system can agree on certain content standards for each subject area [23]. Assessment is another critical aspect of education, where educators gather, analyze, and interpret data on student performance to make informed decisions and evaluate instruction [24]. This process involves tracking progress, motivating students, evaluating instructional approaches, and grading students based on their overall performance or expected competencies [11]. Teaching, in contrast, involves guiding students in developing the required knowledge and skills through instruction [25]. It encompasses design, content selection, delivery, assessment, and reflection, and can be implicitly represented in argument discussions of alignment between course objectives and assessment strategies [7].

As a result, the primary standards-based scientific education reforms prioritize the alignment between the course objectives and assessments, despite the fact that there are other types of alignment between educational components [7]. In light of theoretical or empirical educational research, this may be connected to the expectation that course objectives and teacher-made tests alignment lead to evidence-based students' learning and professional development [24]. This suggests that empirical research or theoretical underpinnings are needed for the educational components to create evidence-based learning output concerning course objectives and teacher-made tests alignment. In view of this, the current study investigated how standards (course objectives) and assessments (teacher-made tests) aligned alongside a theoretical model (in this case, the combination of Porter [26] and Bloom's revised Taxonomy), as depicted in Fig. 1.

Bloom's Taxonomy is a hierarchical model that categorizes learning objectives into different levels of complexity. The levels range from basic knowledge and understanding to more intricate evaluation and creation [27]. The original model, established by Bloom and his colleagues, included six critical elements: knowledge, comprehension, application, analysis, synthesis, and evaluation. However, the original taxonomy faced criticism regarding its development and application, leading to its evolution into the revised Bloom's taxonomy (Booker, 2007; Chyung & Stepich, 2003). Critics argued that the model's emphasis on analytic learning and assessment was not holistic and reduced the acquisition of lower-order learning, such as facts. Consequently, the revised Bloom's taxonomy was developed to address these concerns and provide a more comprehensive framework for educational purposes [7,27]. This revised model consists of six levels of cognitive complexities: *remember, understand, apply, analyze, evaluate, and create*. It is applicable across all

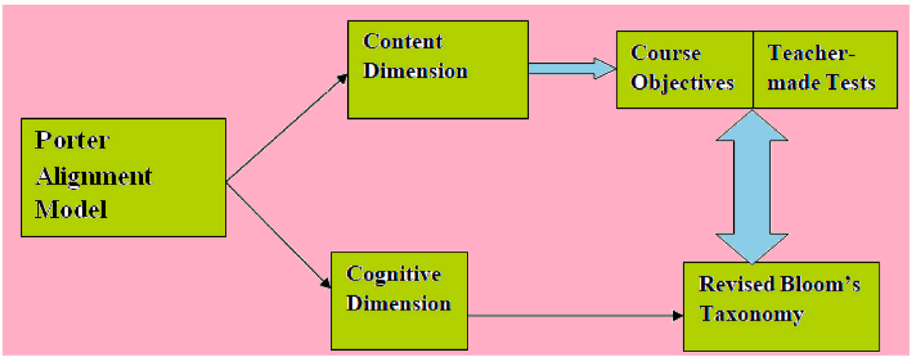

Fig. 1. Demonstrates the combination of Porter's model and Bloom's Taxonomy as a conceptual framework to guide the alignment study.

levels of education and subjects and aims to align assessment with objectives.

The Porter model, also known as Surveys of Enacted Curriculum (SEC), is a tool designed to compare different components of an educational system, such as teaching and assessments, assessments and standards, and standards and teaching in science and math [26]. It consists of two dimensions: cognitive and content. The content dimension comprises performance standards for students or a list of science and math topics without a hierarchy. For this study, topics (key concepts under each unit) from science subject units (collection of different topics) covered in the given semester were used. The cognitive dimension consists of a hierarchical list of cognitive complexities ranging from low to high levels. The revised Bloom's taxonomy was used for this dimension. The course objectives and teacher-made tests were evaluated based on the six levels of Bloom's taxonomy. Using these two dimensions, the model determined how well the test aligned with the course objectives and guided data collection, coding, and analysis (refer to Fig. 1).

3. Method

This section describes the research method, the research settings, participants, data collecting, and data analysis procedures.

3.1. Research method

A research paradigm is a perspective based on a set of assumptions, concepts, and values, about doing research (Stringer, 2007). Besides, the approaches adopted and methods of data collection to be selected depend on the nature of the inquiry and the type of information required (Bell & Waters, 2018). This section will discuss in detail the method used in the research, which employed the Revised Bloom Taxonomy Anderson and Krathwohl [27] and Porter's [26] alignment approach to determine the degree to which teacher–made tests align with course objectives.

As it was discussed in the conceptual framework section, the combinations of frameworks developed by Bloom and Porter categorize both content and cognitive complexities explicitly and enable one to analyze the alignment between different constructs. Bloom's and Porter's taxonomy are the most appropriate models to determine the alignment between assessment and standards [7,28,29]. Hence, for this study, the first dimension involves the science topics covered by a single semester of the high school calendar. The science subjects (biology, chemistry, and physics) are supposed to cover the three last chapters in the second semester of the school calendar. The reason for selecting the science subjects is because the student's scores on the subjects have never exceeded 35 % across all national learning assessments to date [17]. The second dimension involves the six levels of revised Bloom's taxonomy. Hence, this combination provides a way to compute an alignment index that describes the agreement between constructs, i.e. objective and test items, using cell values [26].

3.2. Research site, participants and sampling techniques

This research investigation was carried out in a school located in the X *(to protect confidentiality)* zone of central Ethiopia Regional State, which was recently established. The researchers chose this location for two reasons: Firstly, it provided them with access to relevant data, as it was situated nearby to the institution where the researchers are working. Secondly, it was one of the zones where the recently introduced national curriculum was being pilot-tested in selected schools. Hence, it was an appropriate site for this investigation.

To determine the alignment between the course objectives and teacher-made tests, the researchers collected data from the Ethiopian national curriculum materials for content objectives. Consent was established with the teachers in charge of each subject and the school administration to collect final exam tests from the school. Final exams were preferred over tests constructed by individual teachers, as the tests are typically developed independently by each teacher for only a specific topic or chapter and are not replicable. On the other hand, final exams are constructed by a group of teachers, reviewed, and refined by all the teachers involved in teaching the subjects. The researchers collected a final exam consisting of 28 question items for physics, 30 for biology, and 30 for chemistry. These were the total number of items constructed for the subjects during the semester that the research was conducted.

3.3. Data collecting procedures and its coding

The Porter model was preferred for this study due to "its relative simplicity in calculation and broad applicability" [29] and allows to one to use educational models such as Revised Bloom's Taxonomy. The Porter's alignment index can be used to analyze the degree of

Table 1
Objective table for the three subjects.

	Biology	Chemistry	Physics
Remember	11	27	24
Understand	16	15	20
Apply	4	17	15
Analyze	11	8	2
Evaluate	3	2	1
Create	1	2	2
Total	46	71	63

match between any two educational variables. It allows for determining the alignment between two variables, in this case, course objectives and teacher-made tests using two tables (Table 1 for curriculum objectives and Table 2 for teacher-made tests).

In order to code the materials (the course objectives and teacher-made tests), we utilized Porter's [26] coding scheme, which incorporated a revised version of Bloom's Taxonomy (which we will explain in later sections). To ensure accuracy, we enlisted the help of three subject matter experts who possess extensive background experience. The first expert has taught college physics and science research project methods, as well as courses on teaching methodology, and has published articles in reputable journals. The second expert is a curriculum specialist who has taught biology, curriculum, and education courses, and has also published articles on research methods. The third expert is a chemistry instructor who has taught chemistry and teaching methodology courses for over 20 years, both at the high school and college levels.

The alignment investigation started with a review of the literature and discussions among experts on commonly used action verbs for science subjects. The objectives were collected from three units (unit 4 to unit 6) of each subject because the final examination of the three subjects is expected to involve these three units from grade 10 science subjects according to the school's calendar. Additionally, the three units of each subject are the units that should be covered during the semester according to the school program. During this time, teachers are supposed to create common items that include all of the content areas.

In Ethiopian schools, there could be more than ten sections in a single grade. As a result, more than three teachers could teach the same subject. The teachers who are teaching the subject together will create the final test items. It is expected that all teachers collaborate, review, and approve a common test item set that will be given to every student who has learned the subject matter. These items act as a model for the national exam that will be given after completing grade 12 and help students prepare for it.

To analyze the alignment between these tests and course objectives, a discussion was held among the experts on how to code and categorize objectives from syllabi materials and test items into six levels of revised Bloom's taxonomy. The following action verbs were derived from literature and used for coding purposes to allow the representation of each item or objective in line with one of the relevant action verbs:

- *Remember:* list, indicate, name, recall, identify, recognize, define, state, tell, and describe
- *Understand:* interpret summarize, explain, paraphrase, differentiate, demonstrate, classify, outline, report, point out, draw, label, and compare.
- *Analyze:* analyze, organize, contrast, distinguish, discuss, examine, find out, explore, devise, collect, and interpret.
- *Evaluate:* evaluate, estimate, debate, judge, defend, criticize, justify, recommend, choose, verify, infer, predict, and argue.
- *Create:* generate, plan, produce, construct, create, design, formulate, imagine, improve, inventpropose, develop, and report

After having a common consensus on the action verbs for each level, the experts began to list objectives from the curriculum materials, identify the key terms or phrases for each item and objective, and then the experts categorize (i) objectives of the three subjects' syllabi materials from the three units and (ii) test items developed to measure the three units of each subject into different cells and six levels (see the steps below). Every objective and test item's keyword or phrase is first identified, then categorized, coded, and assigned by each expert into one or more of the cells that determine the intersection of cognitive demand and topics. The data was coded using letters like O and T to stand for objectives and tests, and B, C, and P to denote the three subjects: Biology, Chemistry, and Physics, respectively [30]. To indicate the units from which the items and objectives were derived as well as the cell in which the item or objective is situated with respect to Bloom's cognitive level, numbers were assigned. For example, a Biology objective with the code "BO-6.4.3" belongs to unit 6, Section 4, and is the third objective in the section.

The coders completed their tasks independently using agreed-upon action verbs on the revised Bloom's taxonomy. The inter-rater reliability was determined using Cohen's Kappa index of reliability. The inter-rater reliability was 0.86 ($n = 46$, $p < 0.000$) for biology, 0.81 ($n = 71$, $p < 0.000$) for chemistry and, 0.89 ($n = 63$, $p < 0.000$) for physics objectives. Likewise, the inter-rater reliability was 0.79 ($n = 30$, $p < 0.000$) for biology, 0.76 ($n = 30$, $p < 0.000$) for chemistry, and, 0.85 ($n = 28$, $p < 0.000$) for physics test items. Hence, there is an acceptable agreement for the experts [31,30].

The experts then discussed some disagreements and differences to resolve the concerns [32] and finally developed common tables for both objectives and test items (see Tables 1 and 2). Two constructs are used for coding in the Porter [26] alignment index. According to Porter [26], to determine the alignment between two constructs, there is a need to create tables for the constructs (both objectives and tests), compute ratios of the constructs, and their absolute discrepancies [27]. To accomplish this, common rating tables

Table 2
Test table for the three subjects.

	Biology			Chemistry				Physics		
	Unit 4	Unit 5	Unit 6	Unit 4	Unit 5	Unit 6	others	Unit 4	Unit 5	Unit 6
Remember	1	16	0	6	2	3	5	2	3	3
Understand	1	9	0	2	1	3	3	2	1	0
Apply	0	0	0	0	0	1	4	5	3	9
Analyze	2	1	0	0	0	0	0	0	0	0
Evaluate	0	0	0	0	0	0	0	0	0	0
Create	0	0	0	0	0	0	0	0	0	0
Total	4	26	0	8	3	7	12	9	7	12

were prepared for analysis, ratios, and their discrepancies were computed, and the alignment of the tests and objectives was computed. All of this is displayed below under steps a, b, c, and d.

3.3.1. Objectives and test item identification

To determine the alignment index, the initial step is to construct tables that display the number of objectives and items representing the constructs. Tables 1 and 2 present a two-dimensional matrix that outlines the objectives and final exam assessments of three different subjects. The objectives were collected from the last three units of grade 10 science subjects, specifically units 4 through 6, as they culminate in the final examination. For biology, the experts agreed upon 11, 16, 4, 11, 3, and 1 objectives for remembering, understanding, applying, analyzing, evaluating, and creating, respectively. For chemistry, the objectives for remembering, understanding, applying, analyzing, evaluating, and creating from unit 4 through unit 6 are 27, 15, 17, 8, 2, and 2, respectively. For physics, the objectives for remembering, understanding, applying, analyzing, evaluating, and creating from unit 4 through unit 6 are 24, 20, 15, 2, 1, and 1, respectively.

According to the agreement of three experts, the number of biology teacher-made test items for remembering, understanding, applying, analyzing, evaluating, and creating in units 4 through 6 is 17, 10, 0, 3, 0, and 0 respectively. For chemistry, the numbers are 16, 9, 5, 0, 0, and 0 respectively, with 12 additional items from other contents categorized as “Others”. In the case of physics, the agreed numbers for the same categories in units 4 through 6 are 8, 3, 17, 0, 0, and 0 respectively. On the other hand, the biology test items did not measure any content standard listed under unit 6, ecological interactions.

3.3.2. Ratio calculations

At the outset, we meticulously planned and organized the two tables - one delineating the objectives and the other enlisting the test items (see Tables 1 and 2). Both tables were meticulously designed using a two-dimensional matrix, where the columns represented the three units of each subject and the rows indicated the levels of cognitive demand. We then proceeded to calculate the ratio of each cell value by dividing the number of objectives or tests under each cell by the total number of objectives or tests, and arranged the results in Table 3. By doing so, we first calculated the distribution of tests and objectives across the six cognitive levels individually, as illustrated in Table 3 before discrepancies between tests and objectives were determined.

3.3.3. Discrepancies calculations

Upon computing the ratio of objectives and tests, the discrepancies between them were calculated for each cell and recorded in Table 4 for biology, chemistry, and physics. The discrepancy between the test and objective of the i_{th} cell was computed using the formula $X_i - Y_i$. The overall absolute discrepancy was then determined by summing the absolute discrepancies across all cells.

3.3.4. Alignment calculations

In summary, the process involved identifying the number of objectives and test items obtained from the curriculum materials and school item bank respectively, calculating the ratio of objectives and tests for each level of cognitive complexity, determining discrepancies for each cognitive level, and ultimately computing the alignment between the objectives and tests using Porter’s alignment index equation [26] shown below.

$$P = 1 - \sum_{i=1}^{n} \left| \frac{X_i - Y_i}{2} \right|$$

P in Ref. [26] indicates Porter’s alignment index and ranges from 0 (refers to no alignment) to 1 (refers to highest alignment), X_i indicates the ratio of standards in the i_{th} cell of curriculum objectives table and Y_i represents the ratio of tests in the i_{th} cell of teacher-made tests table, both are with a value from 0 to 1.

3.3.4.1. An example of how objectives were categorized and coded. One of the objectives in the biology grade 10 curriculum course objectives, under unit 4 is, “At the end of this section, the student will be able to define cell cycle”. The action verb used in this objective is “define,” and it falls under the “remember” level of the revised Bloom’s cognitive level. This objective requires students to remember the definition stated in the textbook. It is categorized under the “Cell division” column and the “Remember” row because this cell is denoted for the first level of cognitive demand (Remember) and in unit four (Cell division). It is coded as BO-4.1.1 (see Table 5). “B" represents biology, “O" represents objective, “4″ stands for the unit where the objective was placed, the first “1″ represents the topic

Table 3
Objectives and tests ratios for natural science subjects.

	Biology		Chemistry		Physics	
	Objectives	Tests	Objectives	Tests	Objectives	Tests
Remember	0.237	0.563	0.38	0.53	0.385	0.291
Understand	0.345	0.333	0.212	0.3	0.313	0.107
Apply	0.087	0	0.236	0.17	0.24	0.61
Analyze	0.243	0.099	0.113	0	0.032	0
Evaluate	0.065	0	0.028	0	0.016	0
Create	0.022	0	0.028	0	0.016	0

Table 4
Discrepancies of objectives and tests for natural science subjects.

	Remember	Understand	Apply	Analyze	Evaluate	Create
Biology	0.434	0.312	0.087	0.19	0.065	0.022
Chemistry	0.383	0.2	0.336	0.113	0.028	0.028
Physics	0.124	0.206	0.37	0.032	0.016	0.016

under which the objective is located, and the last "1″ stands for the objective number (in this case, the first objective under the mentioned topic).

Another example from Unit 6 biology is, "At the end of this section, the student will be able to construct food chains and food webs based on observed feeding relationships among common organisms in their locality." The action verb used in this objective is "construct," and students are expected to construct the food chain and food web by analyzing to create a link between the feeding type with the locally available organisms involved in the food chain in terms of producers, primary consumers, secondary consumers, and so on. It is categorized under the row of unit 6, "ecological interactions," column of "Analyze," and coded as BO-6.1.2. "6″ stands for the unit where the objective was placed, "1″ represents the topic under which the objective is located, and the last "2″ stands for the objective number (in this case, the second objective under the mentioned topic of the textbook).

The same procedure was followed to categorize and code the remaining objectives of biology, chemistry, and physics as well.

3.3.4.2. An example of how tests were categorized and coded. One of the physics items constructed by the school teacher was presented as follows. A school teacher created a physics problem involving two charges of equal magnitude but opposite sign, $Q1 = 9 \times 10^6$C and $Q2 = -9x10^6$C, placed 60 cm apart. The question requires students to use the electric field concept and the formula taught in class to calculate the net electric field midway between the two charges, using the value $K = 9 \times 10^9\ Nm^2/C^2$. The question uses the action verb "calculate", which is classified under "apply" in Bloom's revised cognitive levels. This question is identified as one of the questions measuring students' application level from the topic of static and current electricity under Unit 4. As a result, it was coded as PT-4-23, where P stands for physics, T for test, 4 for the unit number, and 23 for the item number (see Table 6).

Another example is the statement "Magnetic field lines never intersect each other," which requires students to write true or false. This question assesses students' ability to recall what they have learned or read in the textbook. The verb used in the question is not explicitly stated, but it requires students to remember what they have already read in their textbook. This verb is classified under "remember" in Bloom's revised cognitive levels. As a result, this question was coded as PT-5-6, where P stands for physics, T for test, 5 for the unit number, and 6 for the item number.

The same categorization and coding procedure was used for all other physics, biology and chemistry questions.

4. Results

The following section pertains to the data analysis and interpretation of research findings to address the fundamental research questions of the study.

4.1. Comparison of teacher made tests against learning objectives of science subjects

Fig. 2 shows the comparison of emphasis placed on course objectives and teacher-made tests about the cognitive demand. As could be seen from Fig. 2, test item construction overemphasized cognitive skill-remember (56 %) because it accounted for more than twice as much emphasis on the biology objectives (24 %) indicated in the policy document. Concerning understanding, the attention given to the objectives (35 %) and teacher-made tests (33 %) shows a fair representation or a closer variation between the two components. A moderate variation is exhibited in the proportion of analysis (24 %) for objectives and (10 %) for the test. For objectives, (8.7 %) apply, (6.5 %) evaluate and (2.2 %) create given emphasis at a decreasing proportion yet they are overlooked during test construction.

According to Fig. 3 above, there appears to be an overemphasis on tests for remembering (53 %) and understanding (30 %), while the course objectives for remembering (38 %) and understanding (21 %) are given less attention. Whereas (17 %) of tests seem to underrepresent (24 %) objectives for the cognitive demand apply and the remaining objectives for analyze (11 %), evaluate (2.8 %), and create (2.8 %) had no representing test item at all. Apart from a heavy emphasis placed on remembering and followed by apply; the objectives are distributed across all the cognitive levels. Instead, the tests focused on lower-order cognitive skills which favor learning by remembering or rote memorization.

Based on the data presented in Fig. 4, it appears that the Physics subject objectives place greater emphasis on remembering (38 %)

Table 5
Demonstrates how objectives were categorized and coded.

Units	Remember	Understand	Apply	Analyze	Evaluate	Create
4: Cell Division	BO-4.1.1					
5: Human biology						
6: ecological interactions				*BO-6.1.2*		

Table 6
Demonstrates how tests were categorized and coded.

Units	Remember	Understand	Apply	Analyze	Evaluate	Create
4. Static and Current …			PT-4-23			
5. Magnetism	PT-5-6					
6. Electromagnetic Waves						

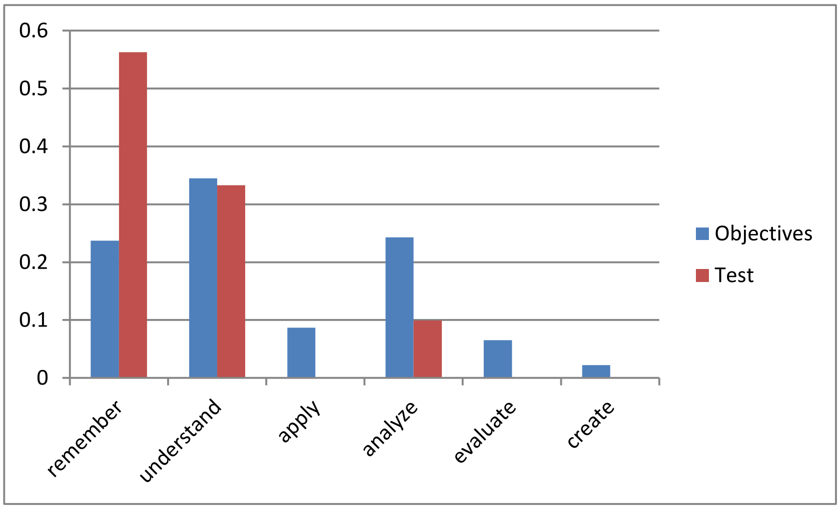

Fig. 2. Comparison of biology learning objectives and teacher made test by cognitive levels.

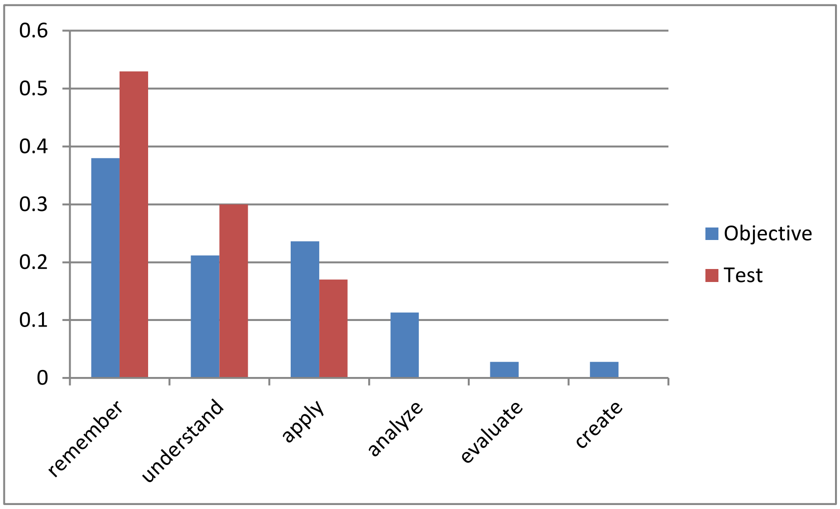

Fig. 3. Comparison of chemistry learning objectives and teacher made test by cognitive levels.

than understanding (24 %) or application (15 %). On the other hand, the higher-order cognitive levels of analyze (3.2 %), evaluate (1.6 %), and create (1.6 %) seem to be given less weight, as they are represented by comparatively small percentages. However, when it comes to testing, the emphasis shifts toward application (61 %) and remembering (29 %), with understanding (11 %) receiving the least emphasis. The remaining cognitive levels are not given any attention in testing.

Tests reflect what has been taught in the classroom which is also dictated by the course objectives stated in the policy document. Thus, other than variations exhibited in the extent of emphasis both the objectives and test item construction placed higher attention to the lower cognitive levels. Again, very high application knowledge demonstrated in the test is not supported by giving adequate emphasis to understanding that gives sufficient conceptual background knowledge for application.

Fig. 5 presents a striking deviation in the application and remembering of scientific knowledge, as shown by a line graph. Specifically, in biology subject, there is a notable difference between the test results and the course objectives. The gap between the two is

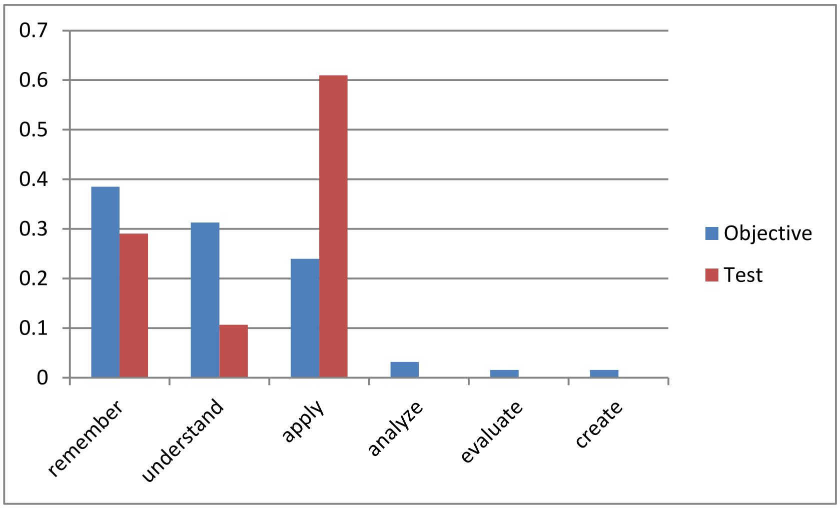

Fig. 4. Comparison of physics learning objectives and teacher made test by cognitive levels.

most evident in the area of remembering (43 %), followed by understanding (31 %), and then further reduced in the area of application (8.7 %). However, the gap widens once again in the area of analysis (19 %). This suggests that biology test items are less aligned with the stated course objectives, which could be due to the curriculum's emphasis on understanding, followed by remembering and analysis. On the other hand, test items that require application, evaluation, and creation have been overlooked, resulting in a lack of emphasis in these cognitive areas.

Even if the trend of the line graph appears similar for Chemistry and Physics subjects and the discrepancy between tests and objectives is not as evident as in biology, misalignment persists in both subjects. Accordingly, Chemistry bears its highest discrepancy on remember (38 %) and apply (33 %) with the subsequent decline from understand (20 %) to the other respective higher order levels. In Physics, on the other hand, a considerable discrepancy between tests and objectives is shown on apply (37 %) along with a substantial decrease from understanding (20 %) to remembering (12.4 %) and the discrepancy becoming much insignificant in the remaining higher levels. Apart from the extent of discrepancy varies from one another, the two most influential discrepancy levels between tests and objectives among science subjects have been exhibited in remembering and understanding Biology, remembering and applying Chemistry, and applying and understanding Physics indicating more significant discrepancy lies between the first three cognitive levels. On the contrary, in the next three higher-order levels the discrepancy between the two components gets narrower and narrower from Biology to Chemistry and then to Physics which demonstrates that tests and objectives are more likely aligned than the lower-level.

According to Porter's alignment index, the alignment between teacher-made tests and objectives is 45 % for Biology, 46 % for Chemistry, and 62 % for Physics. As a result, the alignment between tests and objectives in Physics (62 %) is relatively better when

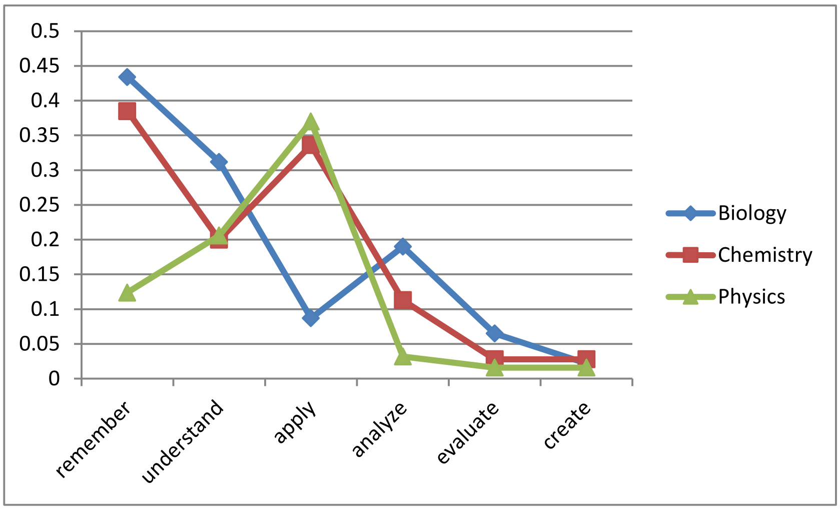

Fig. 5. Discrepancies between objectives and teacher made tests among the science subjects.

compared to Chemistry and Biology, though it may still not meet the necessary threshold.

4.2. Comparisons of learning objectives by cognitive levels across the three subjects

According to Fig. 6, the natural science subjects place a significant emphasis on the first three lower-order cognitive levels, with some variations. Chemistry and Physics objectives appear to receive similar emphasis and are closely linked. For example, both subjects prioritize cognitive demand in the areas of remembering (38 % and 39 %), understanding (21 % and 31 %), and applying (24 % and 24 %) respectively. Biology objectives, on the other hand, prioritize understanding (35 %) and give slightly less attention to remembering and analyzing (24 % each), while applying (8.7 %) receive less emphasis compared to the other two subjects.

4.3. Comparison of teacher made tests by cognitive levels across the three subjects

There is a notable distinction between the values of test items in Fig. 7 and the objectives outlined in Fig. 6, which are based on cognitive demands. Tests should accurately measure the objectives stated in the curriculum document and should also take into account theoretical models about students' competence-based performances. For example, the "remember" cognitive level has the highest proportion of test items in Biology (56 %) and Chemistry (53 %), while "understand" consists of only (33 %) and (30 %) of test items, respectively. In Chemistry, "application" is less emphasized and is completely ignored in Biology, as it seems to be replaced by "analysis." In Physics, however, an over-emphasis is placed on "applying" (61 %), while "remember" (29 %) and "understand" (11 %) take up the remaining share.

On the other hand, none of the three subjects have a single item that represents all three of the higher cognitive levels, except for Biology, which has a proportion of roughly 10 % for the level of analysis. Despite variations in emphasis across subject areas, objectives are distributed across all cognitive levels, but test items are largely concentrated in the first three lower cognitive levels. This highlights an over-emphasis on lower cognitive levels and a lack of attention to the higher levels of test item construction.

5. Discussion and conclusion

5.1. Discussion

The basic assumption of a standards-based approach to science education is the alignment among instruction, assessment, and content standards to create equal opportunities for students to achieve expected learning outcomes [24]. Thus this study intends to analyze the alignment of teacher-made tests with the course objectives of science subjects. Since tests are designed to measure the objectives stated in the policy document, it would be logical to compare how the objectives of science subjects are emphasized and displayed across cognitive levels concerning tests.

Even though both teacher-made tests and science course objectives have a major focus on lower levels in common, the results of the current study indicated that the distribution of test items did not align with the stated objectives. This can be seen in the discrepancies between the test and the objectives. For example, in biology, remembering 0.434 (43.4 %) and understanding 0.312 (31.2 %) constituted the highest discrepancy 0.746 (74.6 %) followed by 0.719 (71.9 %) which is attributed to remembering 38.3 % and apply 33.6 % in Chemistry (see Table 4). In Physics, on the other hand, the highest discrepancy between tests and objectives reaches its highest at 57.6 % by adding up apply 0.37 (37.0 %) and understand 0.206 (20.6 %). The highest discrepancy demonstrated in biology

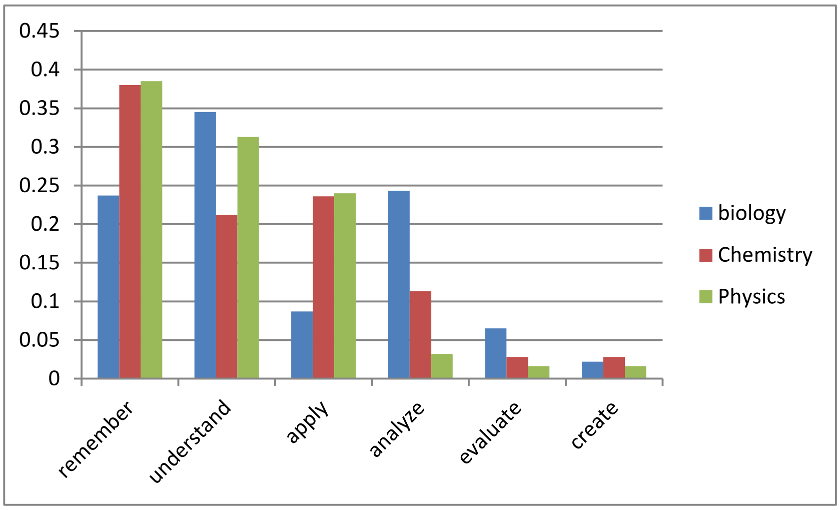

Fig. 6. Comparison of learning objectives by cognitive levels across three subjects.

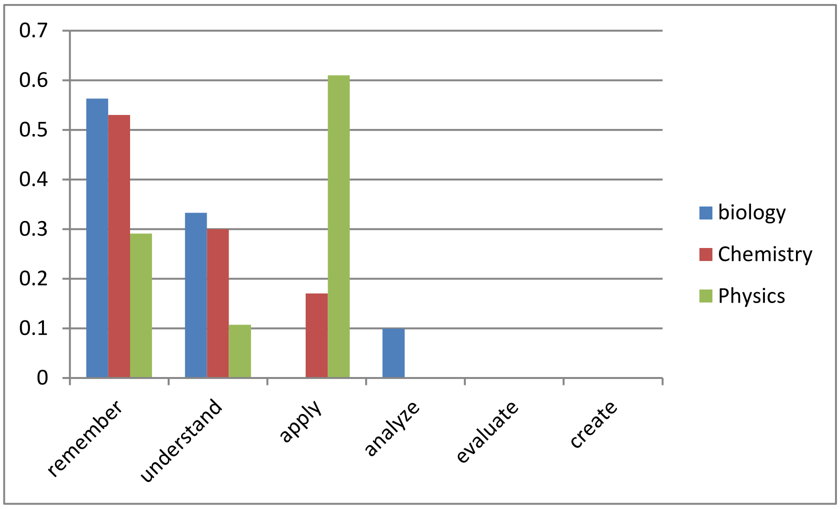

Fig. 7. Comparison of teacher made tests by cognitive levels across the three subjects.

reveals that the tests are the least aligned with the objectives in comparison to Chemistry and Physics.

When examining learning objectives across three subjects, the focus was on the lower-order thinking level (as shown in Fig. 6). Chemistry and Physics placed significant emphasis on the least cognitive level, remembering (38 % and 39 %, respectively). In contrast, Biology placed comparable emphasis on the next cognitive level, understanding (35 %), and then on the higher-order cognitive skill of analysis (24 %). This is in line with previous studies that reveal most assessment items in education focus only on low-level cognitive skills [33–36]. The Ethiopian policy document, however, describes the new curriculum framework as being based on competency-based education, which prepares learners for the challenges of the 21st century and equips them with the necessary coping skills. To achieve this, there needs to be a shift in relative emphasis from lower-order thinking skills to higher-order cognitive demands.

Though the Porter model does not provide an explicit acceptable range for the alignment index, some literature suggests an average acceptable degree of alignment. For instance, some studies indicate that an acceptable degree of alignment between objectives and assessments is at least 70 % and that an alignment of between 60 % and 70 % is also seen as a weakly acceptable level (Webb, 2002), others suggest that more than 75 % response rate was considered as a good indicator of alignment between any two educational components [6,26]. Some others also reported on the work of Porter [26] and computed an inter-rater reliability coefficient of 0.7 for two raters and an inter-rater reliability coefficient of 0.8 for four raters considered an acceptable range [30]. In this respect, though Physics (62 %) had a comparatively better alignment of tests and objectives than Chemistry (46 %) and Biology (45 %), the three subjects did not exhibit the high level of alignment as suggested by the literature.

This generally suggests that the procedures used by teachers to create items did not adhere to a standard-based assessment methodology which demands a checklist for the match between learning objectives and items, item development procedures, assessment blueprint, and performance-level descriptors [37]. According to Alade and Igbinosa [38], there is frequently a genuine discrepancy between the content covered in class and the assessment at the end of the unit test. Additionally, a test that lacks coherence makes it impossible for teachers to make an accurate assessment of a student's academic performance. Carpenter et al. [39] also pointed out that as teachers are the ones who implement the curriculum, they have an enormous impact on whether it succeeds or fails. If the teacher who is supposed to facilitate knowledge to students is incompetent, inexperienced, and unqualified, it will affect the performance of students.

As Carpenter et al. [39] pointed out alignment may vary from subject to subject, as well as from year to year. It can be again argued that changes in test question formats, such as including a performance assessment component to supplement a paper-and-pencil test, may also change the alignment between the standardized test and the content standard. Besides, the below-average alignment of tests and objectives evident in Biology and Chemistry may be related to the strategy adopted by the teachers during test item construction. For instance, the collected data analysis indicated that Biology lacked any representing test item for 15 % of objectives distributed at varying cognitive levels, particularly in unit 6 (see Table 2). Chemistry, on the other hand, incorporated 12 test item questions that don't directly represent the expected content areas (see Table 2). It seems that teacher-made tests suffer from content validity which is largely a function of the content to which test items are a truly representative sample of the content and skills to be learned [40].

According to Gronlund [41], the practical application of a table of specification is to assist classroom teachers in creating well-aligned summative assessments that reflect the subject matter studied and cognitive processes utilized during instruction. Asuru [19] similarly views a test blueprint or table of specifications as a grid displaying content, objectives, and the proportion of questions allocated to each behavioral objective and content topic, all of which aid teachers in testing students across all cognitive domains. The present study's findings reveal that underrepresentation and de-prioritization of certain cognitive skills in test construction may be attributable to a failure to fully utilize the test blueprint or table of specifications.

Furthermore, it has been stated that the updated curriculum framework is connected with competency-based education, which is anticipated to be more rigorous. However, the objectives of the science subjects seem to focus more on lower-order cognitive levels rather than higher ones. For example, remembering and understanding levels combined to account for 27 (58.70 %), 42 (59.15 %), and 44 (69.80 %) of the overall objectives in Biology, Chemistry, and Physics, respectively. Consequently, the majority of science objectives are comprised of lower-level cognitive skills. The literature suggests that this calls for a reconsideration of the distribution of course objectives to align with the aspirations of science subjects [13,42]. This indicates that the content distribution of the newly established curriculum requires reevaluation to incorporate theoretical learning models and high-level student learning. Thus, it is necessary to consider a model-based standards approach in developing or revising the new curriculum, which necessitates higher-order objectives in curricular materials [37].

Concerning item construction in relation to considering theoretical models, although the extent of attention drawn by each subject varies, Chemistry and Physics test items totally (100 %) focused on the first three cognitive levels by neglecting the higher levels. Likewise, in Biology, 89 % of test items were found to consist of remembering and understanding; and 10 % of analysis by simply ignoring apply, evaluate, and create. Thus both tests and objectives placed higher emphasis on the lower levels; however, the apparent difference exhibited on tests is either over or under-representation of test items against the stated objectives at a desired cognitive level. The results from this study show that item construction doesn't demonstrate a model-based assessment approach in which assessments should consider various levels of students' learning and demand a match between test developments with lower and higher-order students' learning [18]. In light of this, Saher et al. [43] argue that assessment strategies that focus predominantly on the recall of knowledge will likely promote superficial learning. This may be related to the fact that teachers are not skillful in constructing test items that fairly represent the objectives across the cognitive levels.

Studies have revealed that merely covering topics is not a reliable indicator of student success on standardized tests. It is the combination of topic coverage and cognitive emphasis that significantly predicts student performance on state standardized tests [44]. Unfortunately, traditional paper-and-pencil test strategies employed by teachers often do not align with the competency-based education outlined in the curriculum. Ahmad et al. [45] and Abate et al. [33] contend that traditional assessments tend to focus on a learner's ability to memorize and recall which are classified as lower-level cognitive skills. While the effectiveness of teacher-made tests depends on their skill in test design, research has shown that many teachers lack this expertise.

Classroom-based performance and portfolio assessments differ from traditional assessments in that they primarily evaluate higher-order thinking skills [34]. Some studies suggest that performance-based assessments are necessary for measuring these skills [12]. To develop higher-order thinking skills, experts suggest that teachers from different subjects and levels of study collaborate and use various strategies [43,46]. Traditional testing, on the other hand, may hinder the assessment of these skills and other crucial 21st-century competencies due to its item format [34].

5.2. Conclusion

Alignment studies offer teacher educators valuable insights into how well teaching, content standards, and assessments are functioning, and enabling informed decision-making. It is worth noting that alignment may differ from one subject to another, and from year to year [29]. Teachers design tests to gauge their students' learning outcomes, and the alignment between these tests and the objectives is crucial to measuring the extent to which students have achieved the standards set. Thus, teachers' evaluations of their students' achievements can be more meaningful when based on this alignment. Based on the research findings presented so far, it can be concluded that the alignment of tests and objectives may vary from one subject to another, especially in science subjects, for various reasons. For example, some teachers lack the necessary techniques and skills to use the table of specifications or test blueprint that would enable them to arrive at a representative sample of instructional objectives and subject matter covered in class. Additionally, traditional assessment strategies, such as paper-and-pencil tests, may suffer from content validity issues when certain units are overlooked, or when test items are included from content not expected to be covered in a given semester. Finally, assessment strategies may not be compatible with competency-based education, which demands alternative assessment methods that can evaluate higher-order thinking skills and other essential competencies required in the 21st century, as noted by Saher [43].

Each traditional and alternative assessment strategy has its own merits and demerits; therefore, teachers should maintain a reasonable balance between the two assessment strategies while designing for test items so that the weakness of one can be compensated with the other. Besides, teachers should be equipped with different assessment strategies and skills so that the table of specifications would receive proper concern during test item construction.

According to the findings of this study, there appears to be a lack of alignment between teacher-made tests and the course objectives of science subjects, with only 45 % of Biology, 46 % of Chemistry, and 62 % of Physics matching the curriculum objectives. This lack of alignment may make it challenging to assess students' higher-order thinking abilities if the curriculum's goal is to develop such skills, but the teacher's test only measures factual knowledge. It suggests that teachers may need to improve their skills in creating tests that align with learning objectives and instructional techniques to accurately determine students' knowledge, skills, and abilities. Therefore, it is recommended that teachers should use assessment procedures and blueprints that consider higher-order cognitive levels and expected learning objectives in the curriculum. Additionally, it is also required to carry out additional investigations to determine whether students have to comply with expected learning course objectives in order to advance to the next grade level.

Data availability statement

Data is included in the article.

CRediT authorship contribution statement

Tsedeke Abate: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ertembo Mishore:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation.

Declaration of competing interest

Authors report financial support, administrative support, travel, and writing assistance were provided by Hossana College of Education. Authors reports that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors sincerely thank the Hossana College of Education, school teachers, and schools who participated in this study.

References

[1] MoE, Curriculum Framework for Ethiopian Education (KG – Grade 12), Addis Ababa: Ministry of Education, 2020 (Ethiopia).
[2] E. Armour-Thomas, E.W. Gordon, Toward an understanding of assessment as a dynamic component of pedagogy. Gordon Commission, Google Scholar, Princeton NJ, 2013.
[3] N.S. Boon, Challenges to curriculum implementation: reducing the gap between the aspired and its implementation through change management, Asia Pac. J. Curriculum Stud. 1 (1) (2018) 14–19.
[4] T.D. Yambi, Teacher training vs. Trainer training in education, Int. J. Res. Innov. Social Sci. (2022).
[5] NRC, A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas, National Academies Press, Washington, DC, 2012.
[6] A. Martone, S.G. Sireci, Evaluating alignment between curriculum, assessment, and instruction, Rev. Educ. Res. 79 (4) (2009) 1332–1361.
[7] G. Näsström, Interpretation of standards with Bloom's revised taxonomy: a comparison of teachers and assessment experts, Int. J. Res. Method Educ. 32 (1) (2009) 39–51.
[8] T. Corcoran, F.A. Mosher, A. Rogat, Learning Progressions in Science: An Evidence-based Approach to Reform, Consortium for Policy Research in Education, Philadelphi, PA, 2009.
[9] W.J. Laidlaw, Teacher-made tests: models to serve specific needs, Clear. House A J. Educ. Strategies, Issues Ideas 39 (6) (1965) 336–339.
[10] T. Teferra, A. Asgedom, J. Oumer, A. Dalelo, B. Assefa, Ethiopian Education Development Roadmap (2018-30). An Integrated Executive Summary, Ministry of Education Strategy Center (ESC) Draft for Discussion: Addis ababa, 2018.
[11] N. Bhaw, J. Kriek, The alignment of the Grade 12 physics examination with the CAPS curriculum:(November 2014–March 2018), S. Afr. J. Educ. 40 (1) (2020).
[12] Y. Abosalem, Assessment techniques and students' higher-order thinking skills, Int. J. Second. Educ. 4 (1) (2016) 1–11.
[13] B. FitzPatrick, J. Hawboldt, D. Doyle, T. Genge, Alignment of learning objectives and assessments in therapeutics courses to foster higher-order thinking, Am. J. Pharmaceut. Educ. 79 (1) (2015) 10.
[14] MoE, Early childhood care and education policy and strategic framework (Ethiopia), Addis Ababa: Ministry Educ. (2010).
[15] National Agency for Examinations (NAE), Ethiopian first national learning assessment of grades 10 and 12 students, Retrieved from, Addis Ababa pdf document (2010), https://www.nae.gov.et.
[16] MoE, Education Statistics Annual Abstract (ESAA), 2003 EC (2010/11 GC), Addis Ababa: Ministry of Education (Ethiopia), 2011.
[17] MoE, Grade 12 Exam Result Press Release, Addis Ababa: Ministry of Education (Ethiopia), 2023. https://moe.gov.et/Publication.
[18] J.S. Bae, D.H. Lee, J.Y. Lee, H. Kim, S.J. Yu, J.H. Lee, B.I. Choi, Assessment of hepatic steatosis by using attenuation imaging: a quantitative, easy-to-perform ultrasound technique, Eur. Radiol. 29 (2019) 6499–6507.
[19] V.A. Asuru, Measurement and Evaluation in Education and Psychology, Port Harcourt: Pearl Publishers International Ltd, 2015.
[20] K.S. Krahenbuhl, Student-centered education and constructivism: challenges, concerns, and clarity for teachers, Clear. House A J. Educ. Strategies, Issues Ideas 89 (3) (2016) 97–105.
[21] A.L. Townley, Teaching and learning science in the 21st century: challenging critical assumptions in post-secondary science, Educ. Sci. 8 (1) (2018) 12. Education Sciences, 8(1), 12.
[22] N.L. Webb, Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education, vol. 6, Research Monograph, 1997.
[23] M.E. Goertz, Standards-based reform: lessons from the past, directions for the future, in: Conference on the Uses of History to Inform and Improve Education Policy, 2007, June.
[24] J.L. Herman, N.M. Webb, S.A. Zuniga, Measurement issues in the alignment of standards and assessments, Appl. Meas. Educ. 20 (1) (2007) 101–126, https://doi.org/10.1080/08957340709336732.
[25] J.M. Shafritz, R.T. Koeppe, E.W. Soper, Facts on File Dictionary of Education, Facts on File, New York, 1988.
[26] A.C. Porter, Measuring the content of instruction: uses in research and practice, Educ. Res. 31 (7) (2002) 3–14.
[27] L.W. Anderson, D.R. Krathwohl, A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition, Addison Wesley Longman, Inc, 2001.
[28] R. Matthews, W.W. Kyi. Alignment between the Learning Objectives and Examinations: A Comparative Analysis of High-School Physics Curricula in Myanmar and South Australia, 2019.
[29] G.W. Fulmer, Estimating critical values for strength of alignment among curriculum, assessments, and instruction, Journal of Educational and Behavioral Statistics 36 (3) (2011) 381–402.
[30] A.C. Porter, M.S. Polikoff, T. Zeidner, J. Smithson, The quality of content analyses of state student achievement tests and content standards, Educ. Meas. 27 (4) (2008) 2–14.
[31] X. Liu, B. Zhang, L.L. Liang, G. Fulmer, B. Kim, H. Yuan, Alignment between the physics content standard and the standardized test: a comparison among the United States-New York State, Singapore, and China-Jiangsu, Sci. Educ. 93 (5) (2009) 777–797.
[32] N.L. Webb, Issues related to judging the alignment of curriculum standards and assessments, Appl. Meas. Educ. 20 (1) (2007) 7–25.
[33] T. Abate, K. Michael, A. Angell, Assessment of scientific reasoning: development and validation of scientific reasoning assessment tool, Eurasia J. Math. Sci. Technol. Educ. 16 (12) (2020) em1927, https://doi.org/10.29333/ejmste/9353.
[34] T. Abate, K. Michael, C. Angell, Upper primary students' views vis-à-vis scientific reasoning progress levels in physics, Eurasia J. Math. Sci. Technol. Educ. 17 (5) (2021) em1958, https://doi.org/10.29333/ejmste/10834.
[35] K. Hussien, T. Nigatie, Assessment of instructional objectives in teacher made examinations at Woldia college of teachers education, Abyssinia J. Business Soc. Sci. 6 (1) (2021) 13–23.

[36] J. Momsen, E. Offerdahl, M. Kryjevskaia, L. Montplaisir, E. Anderson, N. Grosz, Using assessments to investigate and compare the nature of learning in undergraduate science courses, CBE-Life Sci. Educ. 12 (2) (2013) 239–249.
[37] E. Forte, Evaluating Alignment in Large-Scale Standards-Based Assessment Systems, Council of Chief State School Officers, 2017.
[38] O.M. Alade, I.V. Omoruyi, Table of specification and its relevance in educational development assessment, Eur. J. Educ. Dev. Psychol. 2 (1) (2014) 1–17.
[39] T.P. Carpenter, E. Fennema, P.L. Peterson, D.A. Carey, Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic, J. Res. Math. Educ. 19 (5) (1988) 385–401.
[40] J.C. Nunnally, I.H. Bernstein, Psychometric Theory, third ed., McGraw-Hill, New York, 1994.
[41] N.E. Gronlund, Assessment of Student Achievement, Peason, Boston, 2006 (8thed).
[42] O. Nessipbayeva, The competencies of the modern teacher, Bulgar. Comp. Educ. Soc. (2012).
[43] A.S. Saher, A.M.J. Ali, D. Amani, F. Najwan, Traditional versus authentic assessments in higher education, Pegem J. Educ. Instruct. 12 (1) (2022) 283–291.
[44] J.V. D'agostino, M.E. Welsh, N.M. Corson, Instructional sensitivity of a state's standards-based assessment, Educ. Assess. 12 (1) (2007) 1–22.
[45] S. Ahmad, N. Sultana, S. Jamil, Behaviorism vs constructivism: a paradigm shift from traditional to alternative assessment techniques, J. Appl. Linguist. Lang. Res. 7 (2) (2020) 19–33.
[46] E. Mishore, T. Abate, Difficulties pre-service science teachers encountered in conducting research projects at teacher education college, Cogent Education 10 (1) (2023) 2196289.