



Alignment of Assessments Items with Curriculum Objectives for Science at Elementary Level

FoqiaAnjum^a, Syed Nasir Hussain^b, Uzma Shahzadi^{c*}

^aCare Foundation Lahore. ^bAssistant Professor, Department of ECE & ETED, Allama Iqbal Open University, Islamabad. ^cDirector Academics, University of Sargodha

*Email: nasir.hussain@aiou.edu.pk

Abstract: The main objective of the study was to analyze alignment of assessments by the Punjab Examination Commission with the National Curriculum objectives for Science at Elementary level. Web Alignment Tool – WAT (version 2) was used to find the alignment, using four areas of alignment i.e., categorical concurrences, depth of knowledge, range of knowledge, and balance representation. Three experts were taken purposively (with at least four years of teaching science at elementary level and holding at least an MPhil degree in Education). They were trained and the researcher worked as the group leader. It was found that 52% SLOs were on DOK level 1, and 26% SLOs were on DOK level 2. Therefore, level 3 and level 4 are not being worked on. Out of 118 SLOs, 81 SLOs were not targeted in PEC test which can be rated very poor alignment between test items and curriculum objectives.

Key words: Alignment, National curriculum, Punjab examination commission, Science grade 8

1. Introduction

The curriculum is a primary and important indicator of quality education because it includes everything to attain the national goal through linking the objectives with teaching content, methodology, learning outcomes, and the assessment. This includes everything that happens under school authority, because of interaction between students and teachers. An educational experience consists of communications among school children and teachers to achieve specific goals (Khan et al., 2017; United Nations Children Fund, 2000). This includes the whole thing that occurs in a school. The term refers to an education group that includes the learner, the teacher, instructional techniques, and experiences. An educational group is comprised of the instructor, the learner, the instructional technique, and the learning outcome.

Curriculum alignment is one of the means of demonstrating or evaluating the link between assessment, objectives of content, and instruction (i.e., curriculum). Students will have the chance to understand and show what they have accomplished if these factors work cohesively to convey a reliable point about what should be taught and evaluated. It has been studied for several years that alignment is not a new phenomenon. The exact alignment exists if the tasks of the assessment are similar to the tasks of the instruction. Currently, alignment studies have a look at the degree to which objectives and assessment deal with equal content material (Webb, 1999). Alignment of the curriculum with assessment is the degree of dexterity is described as curriculum objectives and assessment are in

settlement and assist in combination with each other to manual the machine towards students getting to know what they are anticipated to recognize and do harmony with each other.

The purpose of exams is to maintain and develop connections between intended curriculum, implemented curriculum, and assessed curriculum. In cases of individual students, the results of examinations are reported according to their results as set by an external group of teachers or examiners and based on a curriculum framework or syllabus (Bhutto & Rind, 2022). The effectiveness of institutions and the public image of institutions are still measured in Pakistan and many other Asian countries by test or examination results. Qualification, classification, and punishment are achieved through examinations, which are surveillance tools. In other words, it enables one to distinguish them and judge them. Thus, it has both intended and unintended consequences (Haertel, 2013).

In determining the aligned expectations and assessments in Mathematics and Science Education (Webb, 1997), the multiple criteria are described as the means for determining the relationship between two or more components. Learning goals have been divided into narrowly defined goals of behavior. For each behavioral goal, domains of all possible test items are defined. Expert panel content assessment remains the main method for assessing aligning the objectives and assessment. According to Mahr (2020), standards-based education, systemic reform, and criterion-referenced tests have led to growing complexity and the need for more systematic measures of alignment.

Any country's economic and social development is heavily influenced by education. But in Pakistan, a low level of skills, productivity, and expectations characterize it (Khan, 2009). The gap between policy making and its implementation in the form of teaching is widening in developing countries (UNESCO, 2020). Implementing these policies in a practical and possibly instructional manner can significantly alleviate these problems. As far as the educational system in Punjab is concerned, it is semi-centralized. The textbooks are created by the Provincial Textbook Boards, but the Punjab Examination Commission (PEC), conducts external examinations as an independent body.

Currently, Punjab Examination Commission (PEC) is responsible for assessing students' learning in Grade VIII. Punjab provincial assembly passed the PEC bill in July 2010 and granted statutory status to the PEC in 2006. In grades V and VIII, the PEC was authorized to conduct high stakes terminal examinations. Grade VIII was examined for the first time in 2008 by the PEC. Public and private schools, respectively, enrolled about a million students in Grade VIII examinations in 2016: Mathematics, English, Urdu, Science, and Islamiyat. In Grade VIII, the PEC conducts an examination that has high stakes (Mahmood, 2013; Matters & Toon, 2012). In addition to mentioning the high stakes of the PEC's grade VIII examinations, Rashid, Awan, Muzaffar, and Butt (2011) reported that schools and teachers are financially rewarded for their scores.

Usually, a curriculum is developed to be observed in practice, providing assessment and evaluation. However, they are only used to improve textbooks in Pakistan (GOP, 2009). Generally, the curriculum has evolved to be observed in practice, provision of assessment, and evaluation. In Pakistan, however, it is exclusively used for improving textbooks (GOP). In the absence of synchronization between instruction and assessment, Student Learning Outcomes (SLO's) cannot be achieved (Smith, 2012). Pakistan's elementary education needs to be improved; the curriculum should be aligned with instruction, resources, and assessment. Thus, in Pakistan, there are many layers of curriculum interpretation that can conflict with the curriculum written as well as affect the curricula taught and learned. (Imran & Wyatt, 2019).

2. Curriculum Alignment Models

Different educators have suggested various models of curriculum alignment. These models are attempts to see the need for congruence between distinct curriculum kinds.

2.1 Leitzel and Vogler's Curriculum Alignment Model

Leitzel and Vogler (1994) it can be argued that alignment happens when content transfer (taught curriculum) and curriculum evaluation (evaluated curriculum) align with the scheduled content (written curriculum). Their standard of aligning curriculum consists of three important phases, namely

- a) Planning to Delivery (PD)
- b) Delivery to Evaluation (DE)
- c) Planning to Evaluation (PE).

Leitzel and Vogler argue that for attaining alignment, preparation, delivery, and assessment unity are crucial. This unit requires that scheduled, instructed, and tested content be integrated. However, ignoring scheduled and tested material is a common practice in traditional training.

2.2 Anderson's Model of Curriculum Alignment

Anderson (2002) indicates that the main element of the curriculum is goals, training, supporting equipment, and evaluation, and claims that the connection between these elements is crucial. Figure 2.6 illustrates this connection. Where A denotes the connection between curriculum and evaluation referring to "content validity," the degree to which exams assess curricular goals. B relates to the level of "content coverage" of curriculum goals in teaching in the classroom. The difference between Relation C and Relation A lies in relation to C's focus on "the possibility of learning," which means whether the teacher is teaching the material.

2.3 Webb's Alignment Model

An alignment model is used by Webb to measure aligned curriculum with norms (written curriculum). This tool has been used by more than 20 countries in the United States to assess the alignment of standards and evaluations for different topics. Alignment is calculated using four criteria according to Webb's model

- a. Categorical concurrence [representation]
- b. Depth of Knowledge consistency [depth]
- c. Range-of-knowledge correspondence [breadth]
- d. Balance of representation.

2.3.1 Categorical Concurrence

Content is of great significance, especially in the topics of science. Unconditional competition measures in the region of content congruence between norms and evaluation. It determines how much of the evaluation covers the national content norms. "The principle of categorical competition among criteria and evaluation are gathered if both papers contain the identical or similar types of material" (Webb, Horton, & O'Neal, 2002).

2.3.2 Depth of Knowledge

At the content level, attempting to measure just the degree of a match is not enough. To be well aligned, it is also vital that the density of content coverage is evaluated with the same norms and criteria that are applied to cognitive demand. The depth of knowledge category indicates that "as long as the students' responses to the assessment are as cognitively challenging as what they are supposed to understand and do as outline in the guidelines" (Webb, Horton, & O'Neal, 2002). Uniformity is obtained if the material is covered with the same cognitive rate by at least 50 percent of items in both papers.

2.3.3 Range-of-Knowledge Correspondence

This group coincides with the scope of understanding and abilities. It determines the extent to which the evaluation includes the expertise and abilities required by the norms. This is achieved by taking into account the number of goals delivered by the evaluation things. To be suitable for this type, every item from a standard must be included in the assessment. It provides equal emphasis to each standard's goals.

2.3.4 Balance of Representation

The emphasis placed on a normal goal in the evaluation is represented by this category. The comparative emphasis of an objective is indicated by the value of this category. It demonstrates the dispersal of items of evaluation. The following formula can be used to determine it calculating the difference in the percentage of goals and the percentage of items linked to that goal in the evaluation.

To use method reviewers (curriculum experts and teachers) from the Web, rate the products in the standards as well as evaluation. First, these assessors are being educated then, on the classifications provided above, they analyze the standards as well as the evaluation and rate each item in the said papers. Reviewers' level of agreement is very essential as it may alter the outcomes of the alignment. Recognizing the absence of agreement among reviewers as an issue, Webb, Herman, and Webb (2006) argue **that** this may be due to the absence of assessors' guidance or uncertainty of goals and standards.

Public examinations in Pakistan support rote memorization and cramming (Rehman, 2004), and these exams do not test students' innovative abilities, nor do they cover major content ideas (Hina, 2008). Many studies have also criticized the authenticity and integrity of public examinations in Pakistan.

Standardized benchmarks are one of the policy actions stipulated in the National Education Policy (2009) (Ministry of Education, 2009). The reason behind these discrepancies may include ignorance of the procedures given in the curriculum while performing the evaluation. It shows a gap between the curriculum and assessment. Educational achievement is determined by evaluating students' progress and implementing instruction at multiple levels of the educational structure. Policymakers do not develop standards for assessing content standards (curriculum frameworks), but they do use the standards to make policy, Researchers can systematically compare the content of the system of education consists of various components to see how well they agree with each other with alignment studies. So, the Standardized curriculum and evaluation strategies locally implemented policies are taken into account. By comparing and judging the agreement of various educational components, alignment studies allow researchers to study their content systematically.

A few studies in Pakistan have already been conducted on this subject due to its importance, but all of them have been focused on the relationship between Curriculum and textbook assessment. Studies especially touching this area of research, i.e. alignment analysis of the National Curriculum with the system of evaluation at the elementary level are almost rare.

The curriculum describes certain skills related to that subject to be mastered by the learners. The curriculum demands certain knowledge, and skills are described in the form of "Students Learning Outcomes". The purpose of assessment is to gather data regarding students' achievement of these outcomes. Assessment of students' achievement of learning outcomes of a particular subject and grade level is impossible unless the assessment items match with the Curriculum. It is therefore important for an assessment tool to be in alignment with intended learning outcomes of the Curriculum.

In this study, the researcher aimed to analyze the alignment of Science of grade VIII used by the Punjab Examination Commission with its prescribed National Curriculum 2006 with the help of the Web Alignment Tool.

2.4 Objectives of the Study

The following objectives were pursued in this study:

1. To determine the alignment of National Curriculum Objectives of Science Grade VIII and assessment items in Punjab Examination Commission exam.
2. To find out the gaps between Curriculum Objectives of Science Grade VIII and assessment items.

The study addressed questions relating to the analysis of curricula alignments. The questions were focused on the alignment between the question papers governed by the Punjab Examination Commission (PEC) and the Curriculum of Science of the elementary level.

To answer the following questions, the present research was conducted:

1. Do the assessments items with the Categorical Concurrence correspond to the primary objectives of the Curriculum?
2. To what extent are the assessment items similar to or different from their respective Objectives in terms of their Depth-of-Knowledge?
3. Are there any items in the assessment which do not target the Balance Representation?
4. Do the assessment items with the Range-of-Knowledge Correspond to the objectives of the Curriculum?

3. Research Methodology

The study followed a quantitative research design. Analysis of the assessment's alignment with Curriculum an objective was conducted through descriptive research. For this purpose, Web Alignment Tool WAT (version2) was utilized. To find out the answer to the research questions of this study, an analysis was conducted using this tool the four criteria of alignment of Objectives of National Curriculum and assessment that include categorical concurrences, depth of knowledge, range of knowledge, and balance representation.

In accordance with Dr. Norman Webb's recommendations, the three content experts were selected. These reviewers included content-area teachers, and the researcher worked as their coordinator or group leader. For this purpose, there were two-day alignment training to analyze eight grade Science Curriculum with the PEC Assessment items, 2021. Content-area teachers having minimum of 4 years teaching experience and hold at least MPhil degree in Education.

3.1 Analysis of the Alignment between Assessment and Curriculum Objectives Using WAT (v2)

To identify alignment between curriculum objectives and assessments, the second version of the Web Alignment

Tool (WATv2) was utilized. More than ten states [of America] have utilized this quantified coding and analysis of objectives and assessments (CCSSO, 2002). The needed approval was gained from the inventor of this tool, Mr. Norman L. Webb, senior research scientist of Emeritus Wisconsin Center for the Education Research University of Wisconsin, Madison.

WAT (Webb Alignment Tool) is a tool that permits national curriculum managers to gauge whether standards and assessments are aligned using an internet application (Webb, 2005). WAT v2 was adapted by the permission of the developer. Use of this tool was to recognize four standards of alignment among the objectives and assessments. Additionally, use of this tool was to create a report that summarized the challenges the assessment items had faced (Webb, 2007).

The four criteria were identified by Webb (2007) as follows:

Categorical concurrence is the measure in the region of content congruence between standards and evaluation. It is compulsory to fulfill the requirement that six assessment items should hit any of the standards. Evaluations are judged by their compliance with state content standards. "The principle of definite competition among standards and evaluation are gathered if both papers contain the identical or similar kinds of material" (Webb, Horton, & O'Neal, 2002).

Depth of knowledge is not enough to determine just the level of competition at the content level. As a prerequisite to a decent alignment, there also needs to be the same standards and evaluations of the complexity of content coverage. This category illustrates that "if what the students are asked on the evaluation is as cognitively demanding as what the standards require them to learn and do" (Webb, Horton, & O'Neill, 2002). Consistency is attained if more than fifty percent of the items in both papers cover the material at the similar cognitive rate.

The Range of Knowledge Correspondence category coincides with the space of understanding and abilities. It measures the extent to which the evaluation includes the expertise and abilities required by the norms. This is achieved by taking into account the total of goals directed by the evaluation items. An agreeable degree of alignment is when at least one item from half of the targets has been included in the evaluation. It provides equivalent weight to each standard's goals.

The balance of the Representation category shows the amount of importance the evaluation places on a normal goal. It is a sign of the relative importance of intent. It demonstrates the distribution of the pieces of evaluation. It can be determined by finding out the difference in the percentage of goals and the percentage of things linked to that goal in evaluation. Achieving this criterion means that items of the same complexity or greater were used to meet more than half of the learning outcomes.

3.2 Data Collection

As per the instruction of Webb (2007), three to four reviewers were needed to ensure reliability. The researcher worked as a coordinator or group leader. It was also required that at least one of the participants should be a teacher having experience in the existing standard-based curriculum. So according to these recommendations, reviewers were teachers that had the experience of teaching the science subjects and had a master level degree in Science.

It has prescribed the role of reviewer in manuals for WAT (v2) Webb (2005), according to the instructions the reviewers must do some tasks which include the following:

- Each standard has a learning goal to measure how well it understands its content;
- It is to determine the extent of understanding degree of every assessment item.
- Identifying the main objective and two secondary objectives for each item
- To determine whether any items are challenged from the wrong source.

Being the group leader or the coordinator, the researcher had to perform a dual role. The major responsibility of the group leader was conducting a two to a three-day training seminar. The main purpose of this seminar was to give the participants first-hand knowledge and understanding about the WAT tool and understanding the levels of DOK. For this purpose, the researcher, as a group leader carried out two-phase steps. Before conducting a meeting with the reviewers, the group leader (the researcher) took the following steps:

Step 1: Registering as a Reviewer

The group leader registered each member of the group (the reviewers) with the Web Alignment Tool by going to the **Login** page.

Step 2: Enter the State Standards

The researcher entered the standards of the National Curriculum in WAT. She entered six Standards and forty-seven Benchmarks to their relative Standards and ninety-nine Students Learning Outcomes to their comparative

Benchmarks.

Step 3: Enter the information of Assessment items

The researcher entered the assessment items in WAT. These items were taken from the paper that was developed by Punjab Examination Commission (PEC) for year 2021 for class VIII. The researcher entered a total of thirty-seven items in a tool which were based on thirty-two MCQ's items, each consisting of three distracters (a, b, c) and five open-ended questions.

Step 4: Create Studies

When the researcher entered all Standards, Benchmarks, Students Learning Outcomes, and assessment items of science subjects in WAT, it was an online process to create studies by pairing the entered objectives with entered assessment items.

Then the researcher took some steps during the alignment seminar that included the following:

Step 5: Training of the Reviewers (Phase I)

The researcher was the group leader, instructing the reviewers on the use of the Web Alignment Tool and training them on the depth-of-knowledge DOK levels. Reviewers were informed that the DOK level of an Objective indicates the complexity and not the difficulty of the objective. The standard information in WAT would be entered, or the already entered information about standards and objectives would be verified to observe which objectives correlate with a certain level of DOK. All of them were assigned codes according to these levels. DOK has four levels which are briefly discussed below:

Table 1: DOK Levels

DOK Levels	Title of the Level
1	Recall
2	Skills and concepts
3	Strategic Thinking
4	Extended Thinking

Level 1 (Recall and Reproduction) Contains facts, definitions, terms, or simple procedures both to be recalled and to be performed. Level 1 consists of students performing rote procedures, following well-known formulas, or carrying out a clearly defined series of steps that can be found in a recipe. There is typically only one step involved in a simple procedure. There is a common understanding that verbs such as "recall," "recognize," "use," "calculate," and "measure" indicate mental work at the recall-reproduction level.

Level 2 (Skill/Concept) Mental processing that extends beyond habitual responses. During Level 2 of the assessment, students must make some decisions regarding the approach they will take to the problem or activity.

Level 3 (Strategic Thinking) Better reasoning, planning, evidence-based reasoning, and higher thinking skills are required than in the previous levels. Level 3 cognitive demands are complex and abstract. There are many answers or that both Levels 1 and 2 are possible that makes it complex, but the multi-step task itself and the more complex reasoning that it requires. A Level 3 question requires students to describe their thinking, while a Level 2 question may only require a very simple explanation or a few words. The level 3 activities are typically those where students have to justify their responses and those with multiple possible answers.

Level 4 (Extended Thinking) Thinking most reasonably and planning over a long period is complex and requires complex reasoning, planning, developing, and thinking.

For the assessors to work on WAT (v2), they had to receive specialized training. A training session was held to enable the reviewers to label the DOK level of each learning outcome/skill and enter it individually in the WAT (v2), as assigning DOK levels both to objectives and to assessment items is critical aspect of alignment analysis.

Step 6: Facilitating the Consensus Process

As a consensus process, reviewers agreed on each Objective's DOK level. After giving the instruction on DOK and the usage of WAT (v2), the team of a group leader entered the DOK levels of each SLO in the tool. A debate was then arranged within the group by the researcher as a leader about the given DOK levels for each learning outcome/skill. They discussed why they allocated the specific DOK level to the SLO. In this step, we were trying to develop a consensus on learning outcomes/skills regarding the DOK levels.

3.3 Training the Reviewers (Phase II) about Coding the Assessment Items

The researcher as a group leader trained the reviewers about the coding process. It was explained to the assessors how to code the objectives (primary/secondary).

Throughout the coaching, the reviewers were up to date that for coding the assessment items. All assessment items were assigned DOK levels, and each assessment item was matched to a standard or objective corresponding to the skill actuality tested (Webb, 2005).

According to WAT (v2) provisions, one assessment item may correspond to up to three objectives, as a result of an assessment item targeting more than one learning skill (Webb, 2005). But the reviewers were told not to code more than one objective to an item unless that item entirely targets each of those objectives. The item should only be coded to Objective A if it targets objective A and to a lesser extent, objective B (Webb, p. 42).

A learning outcome or skill may not directly correspond to an item; however, they may be matched and allocated to any target or standard. The item will be marked Uncodeable, however, if it fails to meet any standards, benchmarks, or learning outcomes.

The information about the assessment was entered into WAT to determine the total items after training and understanding all the criteria of DOK levels. The items of assessment were figured out with their respective DOK level. Each assessment item was assigned a code according to the levels of DOK. If the assessment item number four hit level 1, it was coded as 1. (The word 'hit' has been used to show when matching occurs).

As part of measuring the condition of assessment items, the assessors' views were also gathered in the form of 'sources of challenge' and notes (if needed).

Reviewers considered an item to contain a cause of a contest if they felt that the item is:

- a) It is possible for a student who knows that the correct material is available to still make a mistake or
- b) The student may answer the question correctly even if he or she isn't familiar with the material.

Notes were designed to provide additional information about the item and its relationship to goals (Webb, 2005).

Step 8: Entering the Group Consensus DOK Values in WAT (v2)

Once the group head got the group consensus DOK values for objectives, she entered them into WAT (v2). It was now an easy task and took very little time.

Step 9: Debriefing Process

Reviewers' opinions were also gathered, through 'sources of challenge' and comments (if needed), as part of determining the quality of assessment items. Such debriefing questions helped the researcher to make a report. During this process, the following questions were discussed:

- A. Are the items for each competency covering the essential topics you expected? If not, which competencies weren't appropriately assessed or which topics weren't adequately covered?
- B. Have the items on each competency covered the major performance (DOK levels) you expected?
In the case of no depth-of-knowledge assessment for a competency, what should be done to improve that assessment process?
- C. How do the competencies and assessments at this grade compare with those in other grades with regard to content and depth of knowledge?
- D. How would you rate the alignment between competencies and assessment?
 - (a) Perfect alignment
 - (b) Acceptable alignment
 - (c) Needs slight improvement
 - (d) Needs major improvement
 - (e) Not aligned in any way?
 - (f) Other comments?

The reviewers had to enter their responses to these queries.

Ensured Teachers are Trained

As a method of ensuring that teachers receive training, the group leader develops self-made items, while the reviewers determine the level of DOK for these items. A consensus was reached on 8 items and 2 items were

discussed to determine the consensus.

4. Data Analysis

For each of the four criteria, results were reported, in addition to the levels of acceptable achievement for each. The Web-based tool was used by reviewers for automating their analyses of the data. Through the report generated on WAT (v2), the alignment between Curriculum Standards/Objectives and assessment items was identified and analyzed in light of the online manual that accompanies this tool.

4.1 Findings

Every assessment is meant to evaluate the mastery of the Curriculum SLOs by students, but it is not possible to cover all the SLO's of the target subject in any assessment plan. As the coordinator of this study, the researcher found out that there are several SLO's in the Curriculum of Science Grade VIII which have no respective benchmarks. In other words, there are various SLO's of the Curriculum of Science that do not belong to any Benchmark.

It was decided to limit this study to the analysis of the alignment of Grade VIII Science assessments with the Standard Based National Curriculum only. It was, however, observed in the light of the analyzing the data that there are certain gaps in the alignment of assessment and the National Curriculum. It is pertinent to discuss those problems before having a look at the outcomes of the study.

It was observed that one of the six Standards of the Curriculum i.e., *Science, Technology, Society and Environment*, and its corresponding benchmarks were not properly discussed in VIII Grade Science. Moreover, the assessment plan and the assessment items were not even properly designed according to the Student Learning Outcomes/Skills given in the Curriculum. It is evident from the data that the assessment items are designed in accordance with Student Learning Outcomes and hence, this Standard has been particularly ignored.

It was also identified that in the National Curriculum of Science Grade VIII, there are 24 Benchmarks of the six Standards which have no corresponding SLO's (Student Learning Outcomes / Skill). It is, therefore, important to review these Benchmarks. It has also been identified that the third Standard "*Earth and Space Science*" has just two Benchmarks out of five which have their corresponding Students Learning Outcomes/ Skills.

The *DOK (Depth-of-knowledge) level of Learning Outcomes/Skills* relates to cognitive complexity (Webb, 1997). So, a *Learning Outcome/Skill* is challenging cognitively in relation to mental processing when its stage matches *DOK level 1*. But 57 *Learning Outcomes/ Skills* of the Elementary Science Curriculum out of 118 relate to the *DOK level 1*, which makes almost less than 50%.

It has also been noticed during the research and it has been enumerated that except for showing a high percentage of *DOK level 1*, the percentage of *Learning Outcomes/Skills* from the depth of knowledge levels for VIII grade Science is decreasing gradually. For *DOK Level 2*, it decreases from 33 to 26, and for *DOK Level 3*, it decreases from 17 to 13 and on the higher *Level DOK 4*, it decreases from 11 to 9. This problem of the curriculum needs to be carefully revisited.

One of the ways to identify whether both the documents i.e., Standards and the Assessment tool belong to the same context are the analysis of the *Categorical Concurrences* shows that three standards out of six falls in this category. It is evident from the data that there is no work on these three Standards. It was further noticed that another criterion *Range of Knowledge* also did not fulfill its requirement because according to Webb, 2007, the span of knowledge corresponding to Standards are associated with Assessment items that represent the span of knowledge. There are only one out of six Standards i.e., Physical Science having only a weak Range of Knowledge. This weakness also seriously needs to be checked because it shows the lack of professionalism on the part of the Assessment item developers. The data seem to suggest as if the item developers rely solely on the textbook content rather than keeping the *Students Learning Outcomes/skills* of the Curriculum in focus.

Alignment between Standards and Assessment

The reason behind the weak or absence of *Range of Knowledge corresponding* criteria could easily be understood by the data found in Table below presents the results of evaluating the assessment items based on Learning Outcomes/Skill, Benchmarks, or Standards.

Table 2: Result of Assessment items of PEC Assessment Tool 2021 Targeting the Learning Outcomes/Skills

Total LO's	SLOs targeted	not targeted	%	Benchmarks targeted	LO's targeted more than once
------------	---------------	--------------	---	---------------------	------------------------------

Class VIII	118	81	68.6	11	9
------------	-----	----	------	----	---

Note LO's = Learning Outcomes/Skills

The above Table 1 suggests that almost 69% of Student Learning Outcomes/ Skills for Grade VIII Science were not addressed by the PEC Assessment Tool 2021. It is quite surprising to see that only 11 Benchmarks and 9 learning Outcomes/skills have been targeted and they have been repeated at least twice or triple of times in the assessment items without any reason. Moreover, it also seems inappropriate to consider that an assessment scheme comprising of only 47 assessment items can completely correspond to 118 Learning Outcomes/ Skills in terms of their comparison.

Balance of Representation is a gauge that is interlinked with the *Range of Knowledge correspondence* as it seems like equal sharing of several hits i.e. the number of times the *Learning Outcomes/Skills* were hit by an assessment item. For this purpose, only those *Learning Outcomes/Skills* are taken into consideration that has received at least one hit (Webb, 2005).

The analysis of the data for class VIII reveals that the index of *Balance of Representation* is good i.e. 0.81, 0.93, and 0.67 for the Standards *Life of Science, Physical Science, and Science, Technology, Society* respectively but there is no correspondence of these Standards found in terms of *Range of Knowledge*.

DOK consistency between class VIII Science Curriculum and the assessment tool 2021. From this table, we see that shows that the DOK consistency is though at an acceptable level, it was that 57% of the Learning Outcomes/Skills are on DOK Level 1. It just means that there is inappropriate work on DOK Level 2, 3, 4, and more work needs to be done on higher-order thinking.

The quantified coding of Curriculum and assessment is based on four criteria already prescribed above which WAT (v2) are and Webb Alignment tool and there is qualitative judgment in the form of *Source of Challenge and Notes*. In this section of the assessment tool, the reviewers entered the information or views about the assessment items in which the students could either respond with having the knowledge of the answer or being unable to give a response to the item due to its complexity despite knowing the answers.

The reviewers pointed out two assessment items which are items # 19 and 24. Two of the reviewers pointed out eleven items in assessment tool 2021, i.e. items # 1, 10, 16, 21, 23, 26, 28, 33, 40, and 45. And one reviewer pointed out 33 assessment items which are items # 2, 3, 4, 5, 6, 7, 11, 12, 13, 14, 15, 17, 18, 20, 22, 32, 34, 45, 36, 39, 41, 42, 43, 44, 46, 47. Only one item is there for which one of the reviewers claimed that it was difficult but even this item is also not related to the higher-order thinking skills. All the items which the reviewers pointed out are simple and very easy to solve.

If the students learn the lesson in the book and cram for assessment, it means they are on *DOK Level 1* which is based on knowing or remembering. This discussion reveals that the aligned material (learning material and assessment) of the elementary Science Standard does not stimulate enough, and following are some observations from the current research that support the need for higher-order thinking skills:

- 57% Learning Outcomes/ Skills are based on DOK Level 1
- There are no suitable tasks for DOK Levels 3 and 4

An important thing that was elicited from the present research was that the assessment scheme focuses on the 8th grade Science Textbook rather than considering the *Learning Outcomes/ Skill* of Curriculum. If the assessment is not challenging for the learners and the learning material does not provide a learning atmosphere and more importantly if the assessment is made based on the lower order thinking skills and higher-order complexity level is ignored and marks are awarded, such an assessment and consequently, the awarded grades may not be considered valid.

An earlier study was conducted by Gulzar and Mahmood (2018) using the same alignment tool for the subject of Mathematics at secondary level, revealed that one of the five standards that were, *Reasoning and Logical Thinking*, and its corresponding benchmarks was not addressed in the secondary school mathematics as no learning outcome/skill was designed for those benchmarks and standards. It was also identified that for 21 *learning outcomes/skills* of the secondary school mathematics curriculum, there was no corresponding benchmark available. To meet the needs of the alignment tool WAT (v2), he added five benchmarks to connect those *learning outcomes/skills* with the standards. It was evident that 204 learning outcomes/skills out of 280, that is, almost 75 % of the secondary mathematics curriculum is of DOK level 1, Apart from showing the high percentage of DOK level 1, it is notable that the percentage of the DOK levels of the *learning outcomes/skills* among the class IX and X is decreasing in case of *learning outcomes/skills* of higher level, that is, for DOK level 2 it is decreased from 29 to 21

and for DOK level 3, from 0.03 to 0.01. And on the contrary, this percentage is increased from 68 to 78 for the *learning outcomes/skills* of DOK level 1. It was also noticeable that the secondary school curriculum for mathematics was completely void of any learning outcomes/skills of DOK level 4.

5. Conclusions

The following conclusions were drawn as a result of the findings:

1. Due to the large percentage of items in the science assessment that are level 1 of DOK, and the rarity of level 3 and 4 items, the assessment items aligned with curriculum objectives. Thus, we can say that there is a lack of alignment between the curriculum and assessment.
2. Assessment items with the same content are aligned with the Curriculum's primary objective. Students are forced to repeat the fact in assessments, and this is the main reason behind this. As part of the PEC assessment of science, most items related to the content of the assessment. Students had to reproduce the facts only.
3. There were a smaller number of items assessing the level of difficulty. In spite of that many of the topics in the curriculum included SLOs in level 3 and level 4 the assessment items only required students to achieve SLOs in level 1 and level 2 of DOK. Curriculum and assessment are misaligned due to this.
4. Since the assessment items were 47 and the SLOs were 118, the assessment items covered all the objectives. The number of assessment items was 47 and the number of SLOs was 118.

5.1 Recommendations

As a result of analyzing the data and the outcomes of the current research study, the following discussion concludes, the researcher puts forward the following recommendations to cope up with the weaknesses and the problems found in case of assessment:

- Alignment between Standards and assessment may be shaped in the educational system. In fact, there is no training on how to conduct standard-based assessments. It is suggested that PEC should arrange continuous training sessions for Standard-based assessment and ensure that each item of assessment may be gauge to what extent the level.
- It is recommended that the percentage of higher *DOK Levels* should be increased and the Curriculum may be reviewed in this respect.

It is suggested to revisit and make the balance between a number of assessment items and the number of SLOs.

Acknowledgement

The researchers are highly indebted to Dr. Norman Webb for his continuous support and guidance during the research process.

References

- Anderson, L. W. (2002). Curriculum alignment: a re-examination. *Theory into Practice*, 41(4), 255-260. <https://www.jstor.org/stable/1477412>
- Bhutto, M. A., & Rind, I. A. (2022). Influence of external examination on the teaching and learning of Mathematics at the secondary level education in Pakistan. *Journal of Education and Educational Development*, 9(1), 89-110. <https://files.eric.ed.gov/fulltext/EJ1347669.pdf>
- GOP. (2009). *National Education Policy 2009*. Islamabad: Ministry of Education.
- Gulzar, K., & Mahmood, N. (2018). Challenges to maintaining alignment between secondary level Mathematics curriculum and assessments in Pakistan. *Journal of Research and Reflections in Education*, 13(2), 231-243. <https://www.ue.edu.pk/jrre/articles/Article%207.pdf>
- Haertel, E. H. (2013). Reliability and Validity of Inferences about Teachers Based on Student Scores. William H. Angoff Memorial Lecture Series. *Educational Testing Service*.
- Hina, N. (2009). *Analysis of last five years (2003-2007) English papers of SSC of Federal Board of Intermediate and Secondary Education Islamabad*. [Master's thesis. International Islamic University, Islamabad].
- Imran, S., & Wyatt, M. (2019). Curriculum-Making and Development in a Pakistani University. *The Qualitative Report*, 24(10), 2506-2519. <https://doi.org/10.46743/2160-3715/2019.3808>
- Khan, A., Khan, D. S., Ul-Islam, S. Z., & Khan, M. (2017). Communication skills of a teacher and its role in the development of the students' academic success. *Journal of Education and Practice*, 8(1), 18-21. <https://files.eric.ed.gov/fulltext/EJ1131770.pdf>

- Khan, S. H. (2009). Making People Employable: Reforming Secondary Education in Pakistan. *The Pakistan Development Review*, 48(4-II), 603-617. Retrieved from <http://pide.org.pk/pdr/index.php/pdr/article/viewFile/2645/2612>.
- Leitzel, T. C., & Vogler, D. E. (1994). *Curriculum alignment: Theory to practice*. U.S. Department of Education [position paper ED 371 812]. Retrieved from <http://files.eric.ed.gov/fulltext/ED371812.pdf>
- Mahr, L. (2020). *Education Reform and the Transition to Standards-Based Grading in the United States*. [Master's thesis, Northwestern College Iowa]. Available on https://nwcommons.nwciowa.edu/cgi/viewcontent.cgi?article=1262&context=education_masters
- Rashid, A., Awan, A., Muzaffar, I., & Butt, S. (2011). *Improving education through large-scale testing? A study on primary and elementary level exam in Punjab*. Lahore: Society for Advancement of Education.
- UNESCO (2020). *Bridging the gap between policy-making and teaching: Inspiration from Wales*. Retrieved from <https://www.iiep.unesco.org/en/bridging-gap-between-policy-making-and-teaching-inspiration-wales-13508?language=en>
- United Nations Children's Fund, (2000). *Curriculum report Card*. New York: United Nations Children's Fund. Retrieved from <http://www.unicef.org/education/files/curricard.PDF>
- Webb, N. L. (1997). Determining alignment of expectations and assessments in Mathematics and Science education. *NISE Brief*, 1(2), 1-8. <https://files.eric.ed.gov/fulltext/ED405190.pdf>
- Webb, N. L. (1999). *Alignment of Science and Mathematics standards and assessments in four states*. Council of Chief State School Officers and National Institute for Science Education, Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.
- Webb, N. L., Horton, M., & O'Neal, S. (2002). An analysis of the alignment between language arts standards and assessments in four states. In *annual meeting of the American Educational Research Association, New Orleans, LA*. Retrieved from <http://facstaff.wceruw.org/normw/AERA%202002/Alignment%20Analysis%20Language%20Arts%20%20Four%20states%2031202.pdf>
- Webb, N., Herman, J., & Webb, N. (2006). Alignment of Mathematics State-Level Standards and Assessments: The Role of Reviewer Agreement. CSE Report 685. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*. Retrieved from <http://www.cse.ucla.edu/products/reports/R685.pdf>