

An Analysis of Alignment between Secondary School Mathematics Standards and the Assessments 2013 and 2014 of the Board of Intermediate and Secondary Education Lahore Punjab, Pakistan

Khurram Gulzar* and Nasir Mahmood**

Abstract

Alignment between standards and assessments is acknowledged to be the key factor in a standard based education system. In Pakistan, standard based curriculum was adopted for secondary classes for the session 2012-13 onward and this study was conducted to explore the alignment between the secondary school mathematics curriculum and the assessments made by the Boards of Intermediate and Secondary Education (BISE), Punjab, Pakistan. For this purpose, a web based application called Webb Alignment Tool (WAT) was used to assess the alignment between the secondary school curriculum for mathematics with Assessment tools 2013 and 2014 on four criteria: *categorical concurrence*, *DOK consistency*, *range of knowledge correspondence* and *balance of representation*. Both the Assessment tools met the criterion of *categorical concurrence*. However, they could not meet the minimum acceptable level for *range of knowledge correspondence*. *DOK consistency* was found acceptable but instead of quality of assessment items it was because 73% of mathematics curriculum was found at DOK level 1, the lowest level.

Keywords: Educational alignment, standard based assessment, outcome based education

* Ph.D Scholar, Institute of Education & Research University of the Punjab, Lahore. Pakistan.

E-mail: khurram.ier@hotmail.com

** Professor, Department of Early Childhood and Elementary Teacher Education, Allama Iqbal Open University Islamabad, Pakistan. Email: mahsir1@yahoo.com

Introduction

Before the education session 2012-13, the national curriculum 2002 was being followed in Pakistan for the secondary school classes (Aga Khan University Examination Board, 2012). The curriculum was revised in 2006 and it was for the first time in the history of Pakistan that expected competencies were specified in terms of observable and measureable standards (Punjab Curriculum and Textbook Board, 2014). For the session 2012-13 this standard based curriculum 2006 was introduced for several secondary school subjects including mathematics (Punjab Curriculum and Textbook Board, 2014).

This study was conducted to find the alignment between the secondary school mathematics curriculum and the corresponding assessments made by Boards of Intermediate and Secondary Education (BISE). It was conducted using second version of Webb Alignment Tool, used in several states of USA (Webb, 2007b). The Webb Alignment Tool (WAT) is an Internet application used to automate the process of assessing the alignment between standards and assessments (Webb, 2005). Including Webb Model of Alignment, there are four model of alignments that were used in different states of America. (Council of Chief State School Officers [CCSSO], 2006)

Here is a brief description of those models:

- “Webb” This model of educational alignment was designed by Norman Webb at Wisconsin Center for Education Research (Univ. of Wisconsin-Madison) in coordination with CCSSO and states.
- “Surveys of Enacted Curriculum” (SEC) model was developed by Andrew Porter, Director of the Wisconsin Center for Education Research, and John Smithson (Univ. of Wisconsin-Madison). This model was also made with assistance of CCSSO and states.
- “Achieve” The Achieve model was developed by an organization called Achieve, Inc., based in Washington, DC.
- “Council for Basic Education” -- The Council for Basic Education (CBE), Washington assisted in implementing No Child Left Behind including alignment analysis. This company provided technical support in this regard.

SEC model is used for comparison of standards across the states, the Achieve model is used for comparison of standards within state, and the CBE model is used to compare a state standard to the national standards. The Webb alignment process sounds more suitable in Pakistani scenario as it is used for measuring the alignment between state standards and assessments (CCSSO, 2002; Webb, 2007b) and in Pakistan, both standards and assessments are made at state and province level, respectively.

Webb Model

This model was prepared by Norman Webb (Webb, 1997, 1999). This process is a combination of qualitative judgment by experts, a quantified scheme of coding and an analysis of standards and assessments (CCSSO, 2002). In order to compare the relationship between assessments and standards the criteria that are identified are *categorical concurrence*, *range of knowledge correspondence*, *balance of representation*, and *depth-of-knowledge (DOK) consistency*. After training the teachers and the content specialists, they are involved in a process of coding in which they code the DOK levels of standard (CCSSO, 2002; Webb, 2007b). This is the process of the determination of DOK as expected by each learning outcome prescribed in the educational standards under consideration. It also includes coding of those *learning outcomes* against four levels of knowledge. Webb named them as *recall*, *skill or concept*, *strategic thinking* and *extended thinking* (CCSSO, 2002). After this, those teachers/content specialists, assign DOK levels to assessment items and the corresponding curriculum learning outcomes/skills (Webb, 2007). The correspondence between standards and assessments is analyzed against four criteria i.e. *categorical concurrence*, *range of knowledge correspondence*, *balance of representation*, and *DOK consistency* (CCSSO, 2002; Webb, 2007b).

Here is a brief description of these four criteria as well as the four levels of knowledge as defined by Webb (2007a):

Categorical concurrence. For a particular standard, six assessment items are assumed to be the minimum required number to ensure that the standard really have categorical concurrence with the assessment. According to Webb (2007a) categorical concurrence is the criterion that is used to analyze how far same categories of content appear in standards and assessments. For a given standard, the said criterion is considered met if five or more assessment items are found targeting that standard. The underlying assumption behind devising this yardstick is the understanding that there should be at least six items “related to a learning goal ... to take appropriate decisions about students' performance pertaining to that learning outcome (Webb, 2007a).

Range-of-knowledge correspondence. This is a comparison of expected range of knowledge of a student with the range of assessment items about that standard. The minimum level is considered achieved for any standard when more than half of the *learning outcomes/skills* falling under the standard are catered by the assessment items (Webb, 2007a).

Balance of representation. This criterion is about the balance of emphasis of the objectives of any standard. This balance is considered achieved if every objective for a standard is targeted by the assessment items. The degree of balance is represented with the help of an index value.

If the index value is 1 then it is the indication of perfect balance and it is accepted to be obtained when corresponding items for a learning outcome are correspondingly distributed among the learning outcomes for a standard. '0' Index values means that some major chunk of the hits are on only one or two of all of the learning outcomes hits (Webb, 2007a, p. 15).

DOK consistency. As evident from the name, this criterion deals with the degree of complexity associated with a particular standard. The purpose of this criterion is to ensure that the corresponding items in the assessment carry the requisite complexity as intended for the standard under observation. This is how Webb (2007a) described it:

This criterion is used to measure whether the knowledge produced by the students through assessment has the same complexity level as what is expected in the standards. The attainment of this criterion means that more than half of *learning outcomes/skills* targeted were found to be hit by the items of same or above complexity.

Like the four criteria about the correspondence between the assessment and standards, it is also imperative to understand the concept of Webb (2007a) about four levels of knowledge. In context of the subject of mathematics, this is how Webb has described them:

Level 1. This level is called the *recall and reproduction* level. Recalling of a fact, a simple procedure etc. or performing a simple procedure comes under this category. It is a sort of knowing or not knowing of any fact, definition etc.

Level 2. It is *skills and concepts* level. A student is considered achieving this level when he/she is engaged in some mental processing that is beyond simple recall of any fact or process. This level is more complex as compared to Level 1.

Level 3. This is called *strategic thinking* level and it demands the skills like planning, reasoning etc. A student concluding a result using an evidence will be considered doing a Level 3 activity. This is a complex level as it includes multistep tasks that require higher order thinking.

Level 4. This level is named as *extended thinking*. It includes the tasks that demand higher cognitive skills. In the tasks of Level 4 a student have to make several connections among different content areas. He has to opt or design an approach among the available alternatives. This level demands complex thinking, skill of designing and planning, and most probably will require an extended period of time as compared to the time required for the tasks of level 3.

Alignment studies conducted using WAT are not meant to verify the general quality of standards or assessments of a state. Instead the sole purpose of these studies is to identify and discuss the degree of alignment between standards and assessments (Webb, 2007b).

Using WAT, first study was conducted in 2003 to gauge alignment between the standards and assessments of mathematics, reading and science and by late 2005, 17 states of USA had used this standardized tool (CCSSO, 2006) for conducting alignment studies that underlines the efficiency of this tool for standards vs assessment alignment studies.

Significance

In standard based education systems, it is imperative to gauge degree of alignment between the standards and the assessments by some impartial body (Case, Jorgensen, & Zucher, 2004; Burkam, 2013) as this practice is recognized to be the central tenet to bring about standard based educational reforms (Webb, 1997; CCSSO, 2006). This study performed the same task by measuring the alignment between the secondary school mathematics curriculum and the assessments 2013 and 2014 that are prepared by the Board of Intermediate and Secondary Education Lahore, Punjab.

Research Objectives

The objective of the study was to identify the gaps (if any) between the curriculum standards and the assessments at secondary level in the province of the Punjab

Research Questions

Research questions of study were:

1. How far are the curriculum standards aligned with the assessments at secondary level for the subject of mathematics?
2. Are the standards and assessment addressing the same content categories?
3. Does the complexity of assessment items match with the corresponding standards?
4. Is the breadth of knowledge expected for standards catered appropriately with sufficient number of assessment items?
5. Are the assessment items pertaining to any standard evenly distributed to relevant learning skills?

Research Methodology

The research design used in this study was descriptive. To gauge the alignment between curriculum standards and the assessments the second version of Web Alignment Tool (Watv2) was used. The necessary permission was obtained from the developer of this tool, Mr. Norman L. Webb, senior research scientist of Emeritus Wisconsin Center for Education Research University of Wisconsin, Madison. To find the answers of the research questions of the study, this tool was used to identify four criteria of alignment between the standards and assessments.

After creating the study on the WAT (v2), the coordinator of study conducted a two day seminar at the Institute of Education and Research in which the reviewers were trained about using WAT (v2) and DOK levels of learning outcomes/skills for mathematics. The purpose of this training was to make the reviewers able to label DOK level of every *learning outcome/skill* and enter it in the WAT (v2) individually.

After getting the training about DOK levels and the use of WAT (v2), every reviewer entered the DOK level for every *learning outcome/skill* individually. After completion of this phase the coordinator of the study conducted a debate about every individual *learning outcome/skill* to develop a consensus about its DOK level of every *learning outcome/skill*. After reaching consensus on the DOK levels for the learning outcomes/skills, the next step for the team of reviewers was to assess the complexity of the state assessment items of the BISE assessments 2013 and 2014 and to match the items with the secondary school mathematics standards (Webb, 2005).

After necessary training, the reviewers assigned the DOK level to every assessment item and determined which *learning outcome/skill* reflects the skill being tested by each assessment item (Webb, 2005). The procedure started with the registration of this study on online application WATv2. It was done by the researcher as group leader. This step was accomplished by the group leader who entered the assessments in WAT (v2) tool. This step required the group leader to have information about total number of assessment items, its sections etc. In the new standard based curriculum scheme the standards are organized in three layers: there are five standards which are subdivided into 35 benchmarks for which 280 corresponding learning outcomes/skills are required to be achieved by every individual (MoE, 2006). Those standard were entered into the WAT (v2) according to the same hierarchy.

The reviewers were trained about registration on WAT (v2) and DOK levels. At this stage, the reviewers were introduced with the concept of DOK levels for the subject of mathematics. Since the reviewers were to work on WAT (v2), therefore they were given necessary training about it. The purpose of this training was to make the reviewers able to label DOK level of every *learning outcome/skill* and enter it in the WAT (v2) individually. After getting the training about DOK levels and the use of WAT (v2), every reviewer entered the DOK level of every *learning outcome/skill* under consideration. The purpose of this step was to develop a consensus on DOK level about every *learning outcome/skill*. The researcher coordinated the activity as group leader and conducted a debate about every individual *learning outcome/skill* to develop a consensus about its DOK level. After reaching consensus on the DOK levels for the *learning outcomes/skills*, the next step for the team of reviewers was to assess the complexity of the state assessment items for the secondary level and to match the items with the state standards that most closely relate to the items (Webb, 2005, p. 40). During the training the reviewers were informed that for coding the assessment items they will assign the DOK level to every assessment item and determine which standard/objective reflects the skill being tested by each assessment item (Webb, 2005). As one assessment item can target more than one learning skills so as per provision of WAT (v2), a reviewer could “identify one assessment item as corresponding to up to three objectives” (Webb, 2005, p. 40). An item may not directly be matching to any *learning outcome/skill* could be compared and assigned to any benchmark or standard. But if an assessment item does not match with any standard, benchmark or *learning outcome/skill*, it will be declared “Uncodeable” (Webb, p. 42).

Reliability among Reviewers

In order to judge the reliability of reviewers' coding WAT (v2) produces intra-class correlation and pairwise comparison reports. Intra-class correlation, that “is calculated according to the method of Shrout & Fleiss (1979)” (Webb, 2005, p.114) and pairwise agreement is calculated by determining “if the two reviewers gave the item the same DOK Level or not [and by dividing] the number of agreeing pairs of reviewers by the total number of pairs of reviewers [and that is averaged] across all the items on the assessment” (Webb, p.115).

For the current study, it is apparent from Table 1 that intra-class coefficient and pair wise comparison is 0.7 or above. So the reliability for the assignment of levels of DOK among the five reviewers is at acceptable level.

Table 1*Intra-class Coefficient and Pairwise Comparison of Coding by Five Reviewers*

Assessment Tool	Class	Intra-class Correlation	Pairwise Comparison	Pairwise benchmark	Pairwise Reporting Category
2013	9	0.7	0.8	0.8	0.9
2014	10	0.7	0.9	0.7	0.8

Sample and Sampling technique

As prescribed by Webb (2005), five reviewers were selected using purposive sampling technique for the study having knowledge of educational assessment and educational standards. The reviewers were M.Phil. Education and four of the five reviewers were secondary level mathematics teachers having master's degree in mathematics. One of the researcher worked as coordinator/reviewer as advised in the manual.

Instrument

This alignment study was conducted using second version of Webb Alignment Tool, used in several states of USA (CCSSO, 2002; Webb, 2007b). "The Webb Alignment Tool is an Internet application that allows state and district program administrators to automate the process of gauging alignment between standards and assessments"(Webb, 2005, p. 7).

Data Collection

The reviewers entered the data in WAT (v2) that was accessible to the research after completing the processes discussed the methodology section.

Data analysis

On the basis of the data entered in WAT (v2), the reports were generated, that is an online standard procedure available in the said application WAT (v2).

Findings of the Study

The findings about the alignment between curriculum standards and the assessments were accumulated through the reports generated with the help of the WAT (v2). Summary of consensus data about DOK values and corresponding percentages of learning outcomes/skills for class IX and X is given in Table 2.

Table 2*Consensus Data about DOK Level of learning outcomes/skills for Secondary School Mathematics*

Class	Level 1 (Recall)	%	Level 2 (Skill/Concept)	%	Level 3 (Strategic Thinking)	%	Level 4 (Extended Thinking)	%	Total
IX	97	68	41	29	4	3	0	0	142
X	107	78	29	21	2	1	0	0	138
Total	204	73	70	25	6	2	0	0	280

Note. DOK = Depth of Knowledge; % = Percentage of learning outcomes/skills at each level

In the Tables 3, 4, and 5, 'YES' means that, on a particular criterion, requisite level was achieved between the learning outcomes/skills and the assessment. 'NO' is the indication the respective criterion was not met over an acceptable level and 'WEAK' indicates that the criterion was nearly met, within a margin that could simply be due to error in the system (Webb, 2005).

The Table 3 shows the summary of the outcomes of the alignment investigation among secondary school curriculum and the BISE Assessments 2013 and 2014. It is evident from this Table that the state of alignment between the curriculum of Class IX with Assessment 2013 and that of the curriculum of Class X with Assessment 2014 is almost identical. The prime area of concern is range of knowledge correspondence that is the comparison between the range of knowledge of a student about a standard, and the range of assessment items about that standard in the assessment tool. This criterion is considered met for a given standard when more than half of the learning outcomes/skills of a standard are targeted by assessment items (Webb, 2007a).

Table 3*Summary of Acceptable Levels on Alignment Criteria for Secondary Level Curriculum of Pakistan with BISE Lahore Assessments 2013 and 2014*

Standards Titles	Alignment Criteria							
	Categorical Concurrence		DOK Consistency		Range of Knowledge		Balance of Representation	
	Class XI	Class X	Class XI	Class X	Class XI	Class X	Class XI	Class X
Numbers and Operations	YES	YES	YES	YES	NO	YES	YES	YES
Algebra	YES	YES	YES	YES	NO	NO	YES	YES
Measurements and Geometry	YES	YES	WEAK	YES	NO	NO	WEAK	WEAK
Information Handling	-	YES	-	YES	-	NO	-	YES

Note. Range of knowledge = Range of Knowledge Correspondence; DOK = Depth of Knowledge

In the Tables 4 and 5, first column contains the names of standards addressed in National Curriculum 2006 in class IX and X, respectively. Second column shows corresponding number of benchmarks and third column shows average number of learning outcomes/skills, targeted by the five reviewers.

Table 4

Range-of-Knowledge Correspondence and Balance of Representation between Class IX National Maths. Curriculum and BISE Lahore Assessment 2013

Standards Titles	Benchmarks #	LOs #	# of Los Hit		% of Total		Range of Know	% Total Hits		Balance Index		Bal of Rep
			M	S.D	M	S.D		M	S.D	M	S.D	
Numbers and Operations	6	53.6	12.4	0.55	23.13	0.5	NO	28	1	0.75	0.02	YES
Algebra	5	47	11.2	0.45	23.83	0.69	NO	23	1	0.78	0.02	YES
Measurements and Geometry	5	49.2	14.4	1.52	29.27	3.05	NO	49	1	0.65	0.03	WEAK
Total	16	149.8	12.7	1.62	25.41	3		33	14	0.73	0.07	

Note. LOs = learning outcomes/skills; Range of Know. = Range of Knowledge Correspondence; Bal of Rep = Balance of representation; M = Mean; S.D.= Standard Deviation

Table 5

Range-of-Knowledge Correspondence and Balance of Representation between Class X National Maths. Curriculum and BISE Lahore Assessment 2014

Standards Title	Benchmarks #	LOs #	# LOs Hit		% of Total		Range of Know.	% Total Hits		Balance Index		Balance of Rep.
			M	S.D	M	S.D		M	S.D	M	S.D	
Numbers and Operations	4	20.8	10.8	0.45	51.95	2.67	YES	26	1	0.8	0.03	YES
Algebra	5	44.6	11.6	0.55	26.01	1.01	NO	29	1	0.75	0.02	YES
Measurements and Geometry	9	67.6	11	1.22	16.26	1.65	NO	36	2	0.64	0.06	WEAK
Information Handling	3	12.2	3.4	0.89	27.69	6.02	NO	9	2	0.83	0.04	YES
Total	21	145.2	9.2	3.88	30.48	15		25	11	0.76	0.08	

Note. LOs = learning outcomes/skills Range of Know. = Range of Knowledge Correspondence; Balance of Rep. = Balance of Representation; M = Mean; S.D. = Standard Deviation

Discussion

The total number of learning outcomes/skills for Class IX and Class X are 142 and 138 (see Table 2) but it is observable from Tables 4 and 5 that total number of learning outcomes/skills for Class IX and X are 149.8 and 145.2, respectively. This indicated an alignment issue: as explained by Webb (2006, p. 8), “If no particular [*learning outcome/skill*] is targeted by a given assessment item, reviewers were instructed to code the item to the reporting category level. This coding to a generic standard or benchmark sometimes indicates that the item is inappropriate for a grade level such as targeting a [*learning outcome/skill*] at another grade level. However, if the item is grade-appropriate, then this situation may indicate that there is a part of the content not expressly or precisely described in the *learning outcome/skill*. These items may highlight areas in the standards that should be changed or made more precise. According to Webb (2005, p.154), “If the number [of learning outcomes/skills] is greater than the actual number in the standard, then at least one reviewer coded an item for the [*learning outcome/skill*] but did not find any [*learning outcome/skill*]... that corresponded to the item”. For Class IX for instance, for the standard, Numbers and Operations, more than one reviewer coded six assessment items against the generic category, that is, benchmark or standard. The sum of those hits by five reviewers is 13 and so the average, that is, 2.6 increases the actual total of learning outcomes/skills of the standard Numbers and Operation, from 51 to 53.6 in the third column of the Table 4.

The columns four and five of Tables 4 and 5 show the mean and standard deviation of number of learning outcomes/skills coded by the reviewers and columns six and seven show the average percent and standard deviation of the total learning outcomes/skills which got at least one assessment item coded against them.

The minimum acceptable level for meeting the criterion of *range of knowledge correspondence* criterion is that 50% of the learning outcomes/skills for a particular standard should be targeted by at least one assessment item but the Table 4 shows that only 23.13%, 23.83% and 29.27% of the standards: Numbers and Operations, Algebra and Measurement and Geometry were respectively targeted by the items of the Assessment Tool 2013.

The columns eight and nine indicate that [assessment items] are distributed among all of the [learning outcomes/skills] up-to to some degree (Webb, p.113). As shown in Tables 4 and 5, *balance index* is acceptable for Numbers and Operations, Algebra but it is weak for the standard Measurement and Geometry.

In Class X, despite acceptable level of categorical concurrence criterion the Range of Knowledge Correspondence criterion was not met for three of the four standards.

The reason behind this lacking is understandable from the Table 5 which shows that for the standard: Numbers and Operation 51.95% of the learning outcomes/skills were targeted by assessment items, but for Algebra, Measurements and Geometry, and Information Handling, respectively, only 26.01%, 16.26% and 27.69% of the learning outcomes/skills were targeted which means for the standards Algebra, Measurements and Geometry, and Information Handling, the minimum required criteria, that is, 50% is not met. In other words, even half of the learning outcomes/skills were not targeted by the assessment items of Assessment Tool 2014.

Conclusion

The results of the study for both class IX and X are almost the same. For instance, it is evident from Table 3 that major area of concern in both the assessments tools was the *range of knowledge correspondence*.

Apparently it seems a classic example of the case discussed by Webb (1997), while explaining the term, *range of knowledge correspondence*. Despite strong *categorical concurrence* between the standards and the BISE Assessment Tools 2013 and 2014, “the span of expected knowledge within categories... [was not found to be] entirely covered by” (p. 17) the BISE Assessment Tools 2013 and 2014.

To obtain optimal alignment between the standards and assessments between standards and assessment, it is necessary that assessment items target only the learning outcomes/skills and ideally there should be equal distribution of those assessment items among the learning outcomes/skills. The Table 6 shows that instead of targeting solely the learning outcomes/skills, several assessment items were found targeting the generic benchmarks or goals and many learning outcomes/skills were targeted more than once for no good reason. And this is one major reason that, in general, criterion of *range of knowledge correspondence* was not met.

Table 6

Summary of Problems due to Misallocated Assessment Items of BISE Assessment Tools

Class	LOs Not Targeted	% of LOs Not Targeted	Standards/Benchmarks Targeted	LOs Targeted more than once
IX	99	69.7	11	10
X	92	66.7	13	11

Note. LOs = learning outcomes/skills

Table 7*List of Items Not Coded against Learning Outcome/Skill (by more than one reviewer)*

Class	Item No.	Total items	Total point value of items	%	Percentage within the assessment tool
IX	1, 4, 5,6, 9, 10, 11, 13, 18, 32, 34, 36, 37, 38, 40, 47, 50	17	30	25.6	24
X	1, 3, 4, 7, 8, 9, 10, 13, 15, 24, 25, 31, 33, 37, 38, 39, 40, 41, 42	19	29	24.7	23

Note. Every BISE question paper for secondary school mathematics comprises 52 items with 117 as total point value; % = percentage of point value of items not coded against learning outcomes/skills

For class IX for instance, it is notable from the Table 6 that almost 70% of the learning outcomes/skills are not addressed by the BISE Assessment Tool 2013. It can also be inferred from Table 7 that 17 items (that weigh 26% of the total point value of the BISE Assessment Tool 2013), targeted either a standard or a bench mark and 10 learning outcomes/skills have been targeted at least twice (see Table 6) without any reason. The pattern is almost same for class X with minor differences. Generic standards and benchmarks are targeted instead of *learning outcome/skill* and some of learning outcomes/skills are targeted more than once which is not suitable for a tool comprising only 52 assessment items that has to address a curriculum comprising 142/138 different learning outcomes/skills. As a result the range of knowledge correspondence criterion is affected.

Another issue that has affected the *range of knowledge correspondence* criterion is the mismatch between the number of learning outcomes/skills and the number of BISE Assessment Tool 2013 and 2014 items. According to Webb (2005, p.112) “Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of standards and a large number of objectives [learning outcomes/skills]”, so despite the acceptable level of *categorical concurrence* between standards and both the assessment tools and it is not possible for an assessment tool of 52 items to fully address a curriculum comprising 142 or 138 learning outcomes/skills.

Balance of representation is a criterion that is interwoven with the *range of knowledge correspondence* as it looks for equal distribution of number of hits i.e. number of times the learning outcomes/skills were hit by assessment items. For this purpose, only those learning outcomes/skills are taken into consideration that have received at least one hit (Webb, 2005). Balance index is at acceptable level for every standard except Measurement and Geometry but, for Class IX for instance, since the *range of knowledge correspondence* criterion is not met, that is, less than one forth (23.13%, 23.83%) of the learning outcomes/skills were targeted (see Table 5) by the assessment items so the acceptable balance index doesn't really makes any difference. Similarly, for class X, the balance index is acceptable for Number and Operations standard only because according

to Table 5 the *range of knowledge correspondence* criterion was met for this standard, that is, more than 50% of the learning outcomes/skills for this standard were targeted by the assessment items of assessment tool 2014. For the rest of the three standards, since *range of knowledge correspondence* was not at acceptable level so high value of balance index for two of those three standards only shows that the targeted learning outcomes/skills were given almost equal number of hits by the assessment items but this doesn't count in terms of quality of the assessment tool because sufficient number of learning outcomes/skills were not targeted by the assessment tool 2014, for those standards.

DOK consistency between secondary school mathematics curriculum and the assessment tools 2013 and 2014 is at acceptable level but as it is evident from Table 2, for class IX, 68% of the learning outcomes/skills are of DOK level 1 and for class X, 78% of the learning outcomes/skills of class X mathematics curriculum are of DOK level 1. So it is not difficult to infer that instead of appropriateness of the DOK level of the items of the BISE Assessment Tools the acceptable level of the *DOK consistency* is more because of the fact that majority of the *learning outcomes/skills are of DOK level 1*. According to Table 3, *DOK consistency* is weak only for *Measurement and Geometry* standard of Class IX and as 29 out of 41 i.e. 71% of *learning outcomes/skills* of DOK level 2 belong to this standard, *Measurement and Geometry*. *DOK Consistency* is found weak when majority of corresponding *learning outcomes/skills* are above DOK level 1. So far as item making for DOK level 1 is concerned, it doesn't require any thinking effort on part of item maker. An item of DOK level 1 is at the lowest DOK level and any item has to be of DOK level 1 no matter made intentionally or unintentionally. So the acceptable level of *DOK consistency* for most of the standards doesn't necessarily indicates the quality of assessment items as every item will be at or above the level of 73% of the secondary school mathematics curriculum, that is, according to the Table 2, is of DOK level 1.

Recommendations

In the light of the outcomes of the study, it is suggested to revise the existing curriculum of mathematics to incorporate *learning outcomes/skills* of DOK level 2 and above keeping in view the current practices (Ontario Ministry of Education, 2005; Singapore Ministry of Education, 2012). An assessment framework should be developed for the paper setters to guide them about making of standards based assessment. Through this framework, the paper setters are made bound to develop the question paper to measure achievement level of prescribed *learning outcomes/skills*. The paper setters should be trained how to cover at least 50% of *learning outcomes/skills* of every standard. The ratio between number of assessment items and corresponding *learning outcomes/skills* should also be revisited. It is also suggested to conduct alignment studies regularly by some third party (Case, Jorgensen, & Zucker, 2004; Nasser, Zaki, Allen, Mula, Mutawaha, Ali, Kerr, 2014) so that the organizations like BISEs give due importance to adopting measures to ensure standards-assessments alignment.

References

- Aga Khan University Examination Board. (2012). *History of Pakistan classes IX-X*. Karachi: Aga Khan University Examination Board
- Burkam, A. S. (2013). *Alignment analysis of the common core state standards integrated pathway: Mathematics II to the common core state standards*. Portland: Walch Education.
- Case, B. J., Jorgensen, M. A., & Zucker, S. (2004). *Alignment in educational assessment*. New Jersey: Pearson Education.
- Council of Chief State School Officers [CCSSO]. (2002). *Models for alignment analysis and assistance to states*. Washington: Council of Chief State School Officers.
- CCSSO (2006). *Aligning assessment to guide the learning of all students*. Retrieved from scholar.google.com.pk/scholar_url?url=http://citeseerx.ist.psu.edu/viewdoc/download%3Fdoi%3D10.1.1.465.7548%26rep%3Drep1%26type%3Dpdf%23page%3D11&hl=en&sa=X&scisig=AAGBfm3YWIqb8MfUP6RnTyJN9YQL6V2Xgg&nossl=1&oi=scholar
- Ministry of Education [MoE]. (2006). *National curriculum for mathematics grades I – XII 2006*. Islamabad: Ministry of Education Government of Pakistan.
- Nasser, R., Zaki, E., Allen, N., Mula, B. A., Mutawaha, F. A., Ali, H. A. B., & Kerr, T. (2014). Alignment of teacher-developed curricula and national standards in Qatar's national education reform. *International Education Studies*, 7(10), 14-24
- Ontario Ministry of Education. (2005). *The Ontario curriculum grade 9 and 10 mathematics revised*. Retrieved from <http://www.edu.gov.on.ca/eng/curriculum/secondary/math910curr.pdf>
- Punjab Curriculum and Textbook Board. (2014). *Curriculum implementation framework Punjab-2014*. Lahore: School Education Department Government of Punjab
- Singapore Ministry of Education. (2012). *Mathematics syllabus secondary one to four express course*. Retrieved from [https://www.moe.gov.sg/docs/default-source/document/education/syllabuses/sciences/files/mathematics-syllabus-sec-1-to-4-express-n\(a\)-course.pdf](https://www.moe.gov.sg/docs/default-source/document/education/syllabuses/sciences/files/mathematics-syllabus-sec-1-to-4-express-n(a)-course.pdf)
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Washington, DC: Council of Chief State School Officers.

- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states*. Washington, DC: National Institute for Science Education University of Wisconsin-Madison.
- Webb, N. L. (2005). *Web alignment tool: Training manual*. Retrieved from Wisconsin Center of Education Research website <http://watv2.wceruw.org/Training%20Manual%202.1%20Draft%20091205.doc>
- Webb, N. L. (2006). *Alignment analysis of reading standards and assessments Illinois grades 3-8*. Illinois: Illinois State Board of Education
- Webb, N. L. (2007a). *Aligning assessments and standards*. Retrieved from The Wisconsin Center for Education Research Website http://www.wcer.wisc.edu/news%20coverStories/aligning_assessments_and_standards.php
- Webb, N. L. (2007b). Issues related to judging the alignment of curriculum. *Applied Measurement In Education*, 20(1), 7-25.