AMERICAN UNIVERSITY OF BEIRUT

INVESTIGATING THE ALIGNMENT BETWEEN THE FOURTH-GRADE LEBANESE SCIENCE CURRICULUM AND CLASSROOM ASSESSMENTS

by
MARIAM FAYEZ BAASSIRI

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Arts
to the Department of Education
of the Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
September 2016

AMERICAN UNIVERSITY OF BEIRUT

INVESTIGATING THE ALIGNMENT BETWEEN THE FOURTH-GRADE LEBANESE SCIENCE CURRICULUM AND CLASSROOM ASSESSMENTS

by

MARIAM FAYEZ BAASSIRI

Approved by:

Dr. Karma El Hassan, Associate Professor Education	Advisor
Dr. Saouma BouJaoude, Professor Education	Member of Committee
Dr. Barend Vlaardingerbroek, Associate Professor Education	Member of Committee

Date of thesis defense: September 22, 2016

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name: Baassiri Mariam Fayez
Last First Middle

☑ Master's Thesis ○ Master's Project ○ Doctoral Dissertation

☑ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes
after : **One ---- year from the date of submission of my thesis, dissertation, or project.**
Two ---- years from the date of submission of my thesis, dissertation, or project.
Three --- years from the date of submission of my thesis, dissertation, or project

4/10/2016

Signature Date

ACKNOWLEDGEMENTS

It was a long journey with great challenges to complete this thesis and I could never have reached this stage without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

Foremost, all praise belong to Allah who has given me the strength and patience to finish this study.

I want to express my sincere gratitude to my advisor, Dr. Karma El-Hassan for her patience, immense knowledge, continuous support, and valuable guidance throughout my study. I appreciate all the times she dedicated for the completion of my Masters.

I also want to thank Dr. Barend Vlaardingerbroek for his assistance, continuous and valuable feedback, and his constructive comments that helped refine the quality of this thesis. I am immensely thankful to Dr. Saouma BouJaoude and Dr. Anies Al-Hroub, for all their time and help they provided at all levels.

Special thanks goes to the staff at the Department of Education and mostly to Mrs. Leila Kabalan for her assistance throughout the six years I have been at AUB. In addition, special thanks to the expert reviewers, Hiba Hamdan and Sara Khoury, and school principals for their support and participation in this research study.

I would like to thank my amazing friends Aya, Dana, Jessica, Lea, Lina, Mira, Nissrene, Noura, Rama, Reem, and Samia whose friendship and support in the difficult times made such a difference throughout this work.

Without the unconditional love and support of my parents throughout my life and education, I would not have been able to accomplish what I have. I am forever grateful for their care and motivation from the very beginning till the accomplishment of this work.

Finally, to Wafic, who went above and beyond to support and motivate me throughout the process, I am truly blessed to have you as my husband.

AN ABSTRACT OF THE THESIS OF

Mariam Fayez Baassiri for Master of Arts
Major: Educational Psychology

Title: Investigating the Alignment Between the Fourth-grade Lebanese Science Curriculum and Classroom Assessments

This study focused on the alignment between summative assessments and the Lebanese curriculum. Specifically, the purpose of this study was to investigate the alignment between the fourth-grade summative classroom assessments and the fourth-grade 'Matter and Energy' unit in the Lebanese science curriculum using: content and cognitive level.

Summative assessments regarding the science unit 'Matter & Energy' were collected from a sample of 17 schools (5 public and 12 private) for analysis. The alignment between the classroom assessments and the Lebanese science curriculum were examined using three instruments: Curriculum Coding Sheet (CCS), Items Coding Sheet (ICS), and Item Analysis Sheet (IAS). The curricular learning objectives were used for matching the test items to the learning objectives in terms of content. The cognitive levels in Bloom's original taxonomy were used to first classify each learning objective to a cognitive level, and then match the cognitive levels of test items to the cognitive level of learning objectives and assess the alignment between them both.

Findings showed moderate content alignment; further analysis revealed that public schools had higher content alignment than private schools, and they also covered a wider range of topics than private schools. In terms of cognitive level, there was a low alignment between the assessments collected in the study and the curriculum; however, there were items that were higher in cognitive level compared to the cognitive level of the learning objective they were matched. Recommendations for curriculum developers and future research are discussed in light of these findings.

CONTENTS

FIGURES

TABLES

CHAPTER I

Introduction

Background

Educational assessment is an integral part of the teaching and learning process. Since assessments are the basis for important educational decisions, they need to be valid, reliable, and fair. According to the American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT, NCME, & NEA, 1990) the "Standards for Teacher Competence in Educational Assessment of Students" states that teachers need to have basic competencies regarding assessments and particularly being proficient in developing assessments that are suitable for instructional decisions. Focusing on this 'Standard' there are several steps in developing assessments and one important step in constructing classroom assessments involves preparing items or tasks that are representative of the domain being assessed, and therefore teachers need to develop specifications to ensure that. Teachers then have to prepare relevant assessment items or tasks that match the intended objectives and this process of matching the items closely to the intended objective is part of the curriculum-assessment alignment.

In order to validate the content of summative classroom assessments, studies can be done to examine the alignment between test items, test specifications (content and cognitive skills), and instruction (Miller, Linn & Gronlund, 2011). If these components are in agreement, it means that students are taught what is specified in the curriculum and that tests are assessing what is taught. A basic requirement for the content validity of results of classroom assessments is that the test items and the curriculum's learning objectives should be aligned (Nitko & Brookhart,

2011). Alignment is also needed to produce accurate measures of student achievement of standards, and to make fair decisions.

To align assessments to the curricular learning objectives, there are three traditional methods for alignment cited in the literature: (a) sequential development, (b) expert review, and (c) document analysis (Webb, 1997). Researchers have developed more sophisticated models of analysis based on these methods and the three commonly used models are the Webb, the Achieve, and the Surveys of Enacted Curriculum (SEC) models. The Webb and Achieve models focus on comparing the content of the learning objectives to the assessment, while the Surveys of Enacted Curriculum (SEC) model includes the instructional dimension as well. What differentiates between these three models is the level of analysis used for alignment evaluation; these will be discussed in more detail in the literature review. A preferred model for determining the degree of alignment has not yet been agreed upon since different models have different foci and each has its strengths and limitations in specific situations (Martone & Sireci, 2009). Nonetheless, all reviewed methodologies start with the basic evaluation of the alignment between curriculum and assessment items based on two criteria: content and cognitive levels (Martone & Sireci, 2009).

Statement of the Problem

As mentioned above, a basic requirement for the validity of results of classroom assessments is that the assessment measured should be aligned with the learning objectives specified in the curriculum. The main goal of the present study is to investigate the alignment between the Lebanese science curriculum and classroom assessments in terms of content and cognitive levels. This study tackles the scarcity of information and gap in the literature regarding alignment using classroom assessments since the studies found used state or national-level

assessments (Martone & Sireci, 2009; Osta, 2007; Resnick, Rothman, Slattery, & Vranek, 2004; Hajo, 2010; Sleiman, 2012; Webb, 1999; Webb, 2002). In addition, it addresses the gap in the literature regarding alignment between assessments and curriculum at the elementary level. The present study investigates the alignment between the official fourth-grade Lebanese science curriculum and classroom assessments in schools that follow the Lebanese curriculum at the elementary level.

Research Questions

For the purpose of this study, the following questions were addressed:

1) To what extent do classroom assessments measure the content of the fourth-grade Lebanese science curriculum?
    a. To what extent do classroom assessments measure the content of the curriculum?
    b. How are the assessment items in classroom assessments distributed among the content topics in the curriculum?

2) To what extent are the cognitive levels of items in classroom assessments aligned with the cognitive levels required by the learning objectives as stated in the fourth-grade Lebanese science curriculum?

Rationale

Two main reasons made the researcher select this research problem. First, a basic requirement for the validity of results of classroom assessments is that the assessments be aligned with the curriculum learning objectives. Since important education decisions are being made based on students' results on summative classroom assessments, there is a need to validate

that the assessments are measuring students' achievement of curricular learning objectives as intended, and an alignment study served that purpose.

A second reason why the researcher conducted this study is because most studies done using the common alignment methods were done in the USA and on a national level, comparing national standards to national examinations (Martone & Sireci, 2009; Resnick et al., 2004; Webb, 1999; Webb, 2002). No alignment studies using common alignment approaches at classroom level were found. Furthermore, only three alignment studies in Lebanon were found **(**Hajo, 2010; Osta, 2007; Sleiman, 2012) and all are studies at the national level comparing the official curriculum to national examinations at the middle and secondary level in chemistry and mathematics. Consequently, this study tackled this scarcity of information and investigated the alignment between the official Lebanese science curriculum and summative classroom assessments at the elementary level in schools that follow the Lebanese curriculum.

Significance

The findings of the study have implications for theory and practice. With regard to theoretical implications, the study addresses a current gap in the literature in terms of conducting an alignment study using classroom assessments and at the elementary level in Lebanon.

In relation to practical implications, the findings of the study may help teachers and curriculum developers. For teachers, the results of the study show which content and cognitive demands are currently emphasized in the classroom assessments and this identifies the gaps that need to be addressed both in assessment and in the curriculum. Accordingly, there will be recommendations for better alignment between teachers' assessment tools and the curriculum. For curriculum developers the results of this study might reveal whether the learning objectives

need to be more clearly written in order to facilitate curriculum-assessment alignment; the study hopefully will better inform curricular change.

CHAPTER II

Literature Review

Introduction

Educational assessment is an integral part of the teaching and learning process. Since assessments are the basis for important educational decisions, they need to be valid, reliable, and fair; and this entails that assessments must be developed in a way that the items match the intended objectives of the curriculum. This process of matching the items closely to the intended objective is part of the curriculum-assessment alignment. The following sections will briefly describe assessments and their functions and how classroom tests and assessments should be constructed. Also, the importance of alignment as a measure of content validity and the different approaches to alignment research will be discussed.

Assessments and Their Function

Classroom tests and assessments play a crucial role in evaluating student learning and achievement. Assessments are also used for important educational decisions. But first what is an assessment? "Assessment is any systematic procedure for collecting information, that can be used to make inferences about the characteristics of people or objects" (Reynolds, Livingston, & Willson, 2009). Assessment goes way beyond just testing, as tests are only one method of assessing; assessments might also include interview, observations, and reviews of historical records (Reynolds et al, 2009).

Assessments, according to their function, are categorized into several types with two main ones: summative and formative. The same assessment can be used as summative or formative; for example a science test could be administered to a class, if its purpose was to assign grades then its function serves summative purposes; however if the test was used for

identifying students' strength and weaknesses or to provide feedback, then it serves formative purposes. Information collected from assessments, can be used to make inferences about the characteristics of people for their application in decisions related to policy, student evaluations, instruction, guidance and counseling, and selection, placement, and classification (Reynolds et al., 2009).

Constructing Classroom Tests and Assessments

At schools teachers allocate a considerable amount of time to assessment-related activities, and since assessments are important there are some fundamental competencies that teachers must master according to the 'Standards for Teacher Competence in Educational Assessment of Students' (AFT, NCME, & NEA, 1990) that was developed by the American Federation of Teachers, National Council on Measurement in Education, and National Education Association. The competencies that are mentioned in the 'Standards' state that teachers need to be skilled in choosing and developing assessment procedures that are appropriate for instructional decisions; administering, scoring, and interpreting results of both externally produced and teacher-made assessment methods; developing valid grading procedures which use students' assessments; communicating assessment results to students, parents, and other educators; recognizing unethical and inappropriate assessment methods and uses of assessment information (AFT, NCME, & NEA, 1990).

Focusing on the first standard mentioned in the 'Standards', teachers are supposed to be proficient in developing assessment methods that are suitable for instructional decisions. There are some steps that teachers could follow when constructing assessments in order to prepare valid and reliable classroom tests and assessments. For constructing classroom tests and assessments, the first step is determining the purpose of the measure, then developing

specifications, then selecting appropriate assessment tasks, and finally preparing relevant assessment tasks (Miller et al., 2011).

The purpose of the measure is best described in terms of its location in the instructional process: before, during, or after instruction (Miller et al., 2011). Purposes for testing done before instruction may be to determine readiness and whether students have the background knowledge and prerequisite skills needed for instruction, or to determine student placement based on the extent students have previously achieved the objectives of a planned instruction (Miller et al., 2011). Purposes for testing done during instruction is mainly formative where assessments are used to monitor learning progress, provide feedback to both students and teachers, and to detect any misconceptions (Miller, et al., 2011). Purposes of testing done at the end of instruction are mainly summative for assigning grades, however, the assessments can also be used for formative reasons such as for giving students feedback, for assigning remedial work, and for assessing instruction (Miller et al., 2011).

After determining the purpose of the assessment, teachers have to develop specifications for it in order to ensure the assessment measures a representative sample of relevant tasks. Building a table of specifications, which is a two-way chart, is one of the tools used for this purpose. It first involves preparing the list of learning objectives; for example, if we were developing a summative test, we can use the curriculum to extract learning objectives that were guiding our instruction. The list of learning objectives has to be limited to ones that can be measured by a classroom assessment, and would describe the type of performance students are expected to show. Next, the instructional content has to be outlined, and this would indicate the content area in which the performance will show (Miller et al., 2011). After preparing the list of instructional objectives and the instructional content, the table of specification could be

constructed where it relates the content to the cognitive levels required in the objectives. The chart also specifies the number of items in the test and allocates them to each objective and each content area (Miller et al., 2011). This process will allow the teacher to align the test items to each objective and content area. As previously mentioned, the table of specification will help ensure that the assessment measures a representative sample of relevant tasks and this is very important in relation to content validity which refers to the degree the sample of assessment tasks is representative and relevant to the domain of content (Nitko & Brookhart, 2011).

The third step in constructing a test or an assessment involves selecting the appropriate type of items and assessment tasks. Classroom tests can consist of objective items, essay items (one form of performance item), and other types of performance tasks. Objective test items may be either supply types that include short answer and completion items, or selection type which includes matching, true-false, and multiple-choice items (Miller et al., 2011). Essay items can be extended-response or restricted-response questions. Extended-response essay questions are best used when assessing complex cognitive objectives that involve application, evaluation, and the ability to select, organize, and integrate ideas, while restricted-response essay questions are used to measure knowledge of factual material. Other types of performance assessments are classified into extended-response and restricted-response performance tasks and these include oral presentations, constructing graphs, use of equipment/instruments...etc. These differ from essay items as they allow students to perform skills in more realistic contexts (Miller et al., 2011). Both objective test items and performance assessment tasks have their own advantages and limitations which make them more appropriate for some purposes while inappropriate for others (Miller et al., 2011).

An important principle in selecting which type of test item or assessment task to use is to "select the one which provides the most direct measure of the intended learning outcome" (Miller et al., 2011). So if the learning objective is writing or listing, then the item type must be a supply type, if the objective involves using lab equipment then performance task is to be used, and if the objective involves identifying a correct answer then a selection-type item should be used (Miller et al., 2011). The final step in constructing a test or an assessment involves preparing relevant assessment items or tasks that match the intended objectives and are a representative sample (Miller et al., 2011). The process of matching the item or task closely to the intended objective is part of curriculum-assessment alignment.

Overview of Alignment

Before administering a test or an assessment, it is important to validate its content and ensure that the assessment items or tasks are aligned with the intended objectives and are ready to be administered. According to the Merriam-Webster dictionary (2015), 'to align' is defined as "to arrange things so that they form a line or are in proper position". Similarly, in educational settings, one of the alignments in assessment may refer to the degree to which learning objectives and assessments are matching or are in agreement. In order to validate the content of summative classroom assessments, alignment studies are done to examine the alignment between test items, test specifications (content and cognitive skills), and instruction (Miller et al., 2011). If these components are in agreement, it means that students are taught what is specified in the curriculum and that tests are assessing what should be taught. Furthermore, alignment studies provide content validity evidence that supports the interpretations of test scores; this will be discussed next.

Alignment as a Measure of Content Validity

The central purpose of any assessment is to measure students' achievement on a set of specified learning targets, then the results are used for instructional and educational purposes. The validity of assessment results determines the quality of inferences and evaluations that were done. Validity refers to the appropriateness and soundness of interpretations and uses of assessment results (Nitko & Brookhart, 2011). A basic requirement for the validity of classroom assessment results is that the assessment measured should be aligned with the curriculum's learning objectives (Nitko & Brookhart, 2011). In addition, assessments should match the scope of content covered in the learning objectives, and should match the depth of thinking and cognitive skills required by the objectives (Nitko & Brookhart, 2011). To increase the validity of interpretations and uses of assessment results, there are few things that need to be considered: (a) the content and specifications of the assessment (content validity); (b) the construct of the characteristics being measured (construct validity); (c) the relation of the results to other significant measures (criterion validity); and (d) the consequences of the uses of assessment results (consequential validity) (Miller et al., 2011). On the other side, according to Miller, Linn & Gronlund (2011) when talking about teacher-made classroom assessments "it is not practical or necessary to have evidence dealing with all four considerations… the primary concern would be content" (p. 75).

Content validity refers to the degree the sample of assessment tasks is representative and relevant to the domain of content (Nitko & Brookhart, 2011). According to Nitko and Brookhart (2011), content relevance is validity evidence that focuses on the extent the assessment tasks are relevant to what is being assessed, while content representativeness focuses on "whether the assessment tasks are a representative sample from a larger domain of performance" (p. 46). In

other words, it means that the tasks should reflect the content and learning objectives specified in the curriculum (Nitko & Brookhart, 2011). For example if a teacher was constructing a spelling test from a given list of 200 words, it might be time-consuming to have all the 200 words included on the test and therefore the teacher must choose a representative sample of 20 from the 200 words in such a way that if a student was able to answer 80% of the words correctly, it is appropriate to infer that the student can answer 80% of the 200 words correctly (Miller et al., 2011). In addition, the tasks should also reflect the level of thinking specified in the objectives; for example if the objectives state that the students should be able to spell the words, it would be more appropriate to use a spelling test where the students actually spell the words instead of identifying the correct spelling from multiple-choice items.

To ensure content representativeness and relevance, teachers must have a description of the learning objectives and the content to be learned, and each test item or assessment task needs to be checked to make sure it matches the important content and learning objectives. Defining the domain to be assessed should be derived from the goals and objectives stated in the curriculum. Furthermore, the type of cognitions required to perform the tasks need to be analyzed to make sure relevant thinking skills and processes are being assessed (Nitko & Brookhart, 2011). A table of specifications described earlier is one tool that can be used to specify what the assessment should measure. According to Martone and Sireci (2009) “although not everything that is listed in the standards or taught to the student can or should be assessed, alignment research can illuminate how much and to what degree the standard coverage or instructional content has been assessed” (p. 1334).

Traditional Methodologies for Alignment

According to Webb (1997), " assessments must achieve a high degree of match between what students are expected to know and what information is gathered on their knowledge" (p.4). In order to determine whether expectations and assessments are aligned, Webb (1997) outlined three methods for alignment: (a) sequential development, (b) expert review, and (c) document analysis.

Sequential development. In this method, standards and assessments are developed in a serial manner where standards are developed first and then based on them test developers construct assessments. This way the standards act as a blueprint for the content and structure of the assessments (Case, Jorgensen, & Zucker, 2004; Roach, Niebling & Kurz, 2008; Webb, 1997).

Expert review. In this approach, a group of content specialists, who are knowledgeable of the curricular content and educational assessment development, analyze the standards and assessments that have been already developed and judge the alignment between them. Educators, administrators, parents and other public figures in addition to content specialists may sometimes do the review process (Case et al., 2004; Roach et al., 2008; Webb, 1997).

Document analysis. In this approach, the documents, which consist of the curricular standards and assessments, are encoded at first for their content and structure. Then the alignment between them is compared after quantifying and analyzing them systematically. Other documents used, such as textbooks, can also be used in this method of alignment. (Case et al., 2004; Roach et al., 2008; Webb, 1997)

For more sophisticated alignment analysis researchers have developed models, which use the methodologies discussed above, and these models are reviewed next.

Models for Alignment

There are several models to assess the degree of alignment between assessments and curriculum. Bhola, Impara, and Buckendahl (2003) classified these models into three types ranging from simple models to more complex models. A simple model form the basis for the rest of the models and the only dimension assessed is the degree of alignment between test items and content standards (Bhola, Impara, & Buckendahl, 2003). In a simple model, each test item is examined by subject-content experts, and by using a Likert Scale they would indicate the alignment degree by matching processes between objectives and items; the scale would range from 'no match at all' to 'match exactly' (Bhola et al., 2003). A moderate complex model would add other dimensions to the content dimension, such as the cognitive complexity dimension. If so, content panelists would examine the match between the objectives and assessment items using both content and cognitive complexity dimensions. Adding more components or dimensions to be assessed, adds more criteria to the alignment research and it is more likely that the amount of items that match the standards will be less (Bhola et al., 2003). Three common models with different levels of analyses for alignment research, the Webb, Achieve, and Surveys of Enacted Curriculum (SEC) model will be discussed next.

The Webb Alignment Model

The Webb methodology is a complex model used for alignment research and it assesses the alignment between curricular content standards and assessments over five different dimensions: (a) content focus; (b) articulation across grades and ages; (c) equity and fairness; (d) pedagogical implications; (e) system applicability (Webb, 1999). However, in alignment studies,

only the first dimension, content focus, is applied because the rest are too complex to be measured. For the content focus dimension, there are four criteria that have to be achieved in order to establish the alignment.

The first criterion is categorical concurrence, which is similar to content validity in the sense that it indicates whether both the curriculum and assessments incorporate the same content (Webb, 1999). This criterion focuses on broad content areas and looks at the total number of item-objective matches within a content domain. Each item-objective match is called a hit, and for successful alignment, six hits are required for each content domain, hence hits must be within the same content domain (Webb, 1999). The total number of item-objective matches or hits is averaged to determine the average number of items assessed per standard (Webb, 1999).

The second criterion measured is depth of knowledge, which focuses on the cognitive demand of specific objectives and it compares them to the cognitive demand of items in the test that are matched to that objective. Webb listed four levels of cognitive areas: recall, skill or concept, strategic thinking, and extended thinking; however, they can be modified for a particular study (Webb, 1999). The criterion is that at least 50% of the items matched to an objective must be at the same cognitive level specified by the objective or higher.

The third criterion, range of knowledge, analyzes the breadth of standards or content domain and compares them to the breadth covered in the assessment. The number of objectives in a content area measured by at least one item is considered, and a minimum of 50% of the objectives needs to be measured by at least one item. This criterion assumes that at least half of the content domain should be assessed and that all objectives have equal weighting.

The fourth criterion, balance of representation, focus on the extent to which the items in the test are distributed uniformly across objectives within a standard in order to represent breadth and depth of standards (Webb, 1999). The criteria focuses on the difference in the proportion of objectives and the proportion of hits assigned to the objective and a balance index is calculated. An index measure near '1' corresponds to a balanced assessment where most of the objectives are measured by the same number of items, and an index measure near '0' corresponds to an unbalanced assessment (Webb, 1999). The Webb model provides the basis for the two models discussed next.

Application of the Webb Model. Webb (1999) applied his methodology across four states in an alignment study between mathematics and science assessments and state standards. Six reviewers worked on comparing the match between mathematical assessment items and standards in order to assess the alignment based on four of Webb's criteria: categorical concurrence, range-of-knowledge consistency, depth-of-knowledge consistency, and balance-of-representation. The process was done across four states in America. Reviewers agreed about the depth-of-knowledge through group discussions, and their decision was used as a baseline to compare the standards and assessment items to the level of cognitive ability intended. Results showed different levels of alignment across grade levels and states. The categorical concurrence criterion had the strongest alignment as three of the four states satisfied this criterion. Items were evenly distributed among objectives and this fulfilled the balance-of-representation criterion. The remaining two criteria had weak alignment as results showed that assessment items generally targeted low cognitive level i.e. low depth-of-knowledge, and items did not cover the range-of-knowledge intended in the standards.

The Achieve Model

The Achieve method is designed to yield both quantitative and qualitative aspects of alignment between standards and assessments (Martone & Sireci, 2009). It measures alignment using six criteria: (a) accuracy of test blueprint; (b) content centrality; (c) performance centrality; (d) challenge; (e) balance; (f) range (Martone & Sireci, 2009). Regarding the first criterion, accuracy of test blueprint, a team of reviewers ensures that every item in the test corresponds to at least one objective within the standards. Content centrality criterion measures the quality of the alignment between the content of the test items and the content of the related objectives. According to Bhola et al. (2003), content centrality is consistent with Webb's categorical concurrence dimension, and performance centrality is consistent with the depth of knowledge dimension. Performance centrality focuses on the cognitive demand of the assessment tasks.

Challenge, the fourth criterion, consists of two factors: source of challenge and level of challenge. Reviewers evaluate the source of challenge by deciding if the source of challenge in answering the question comes from the content of the item or from other factors such as the language of the item that might provide a clue, or other technical issues (Bhola et al., 2003). Reviewers also evaluate the level of challenge by deciding whether the items in the assessment cover an appropriate range of difficulty in the grade level the assessment is aimed for (Bhola et al., 2003). The balance criterion is a holistic evaluation where it evaluates the extent the content and skills required by test items are given equal weight and importance to the ones in the standards. The range criterion, which is a quantitative measure, measures the standard coverage by determining the extent the content and skills assessed are a representative sample from the standards' content domain (Bhola et al., 2003).

Application of the Achieve Model. Resnick and colleagues (2004) applied the Achieve methodology across five states in America in order to evaluate alignment of assessments to state standards in English language arts (ELA) and Math. Expert reviewers worked on comparing the match between assessment items and standards in order to assess the alignment based on six criteria: accuracy of test blueprint, content centrality, performance centrality, challenge, balance, and range. Expert reviewers who represented classroom teachers, curriculum specialists and content experts were trained to illustrate the rating criteria in this method. Results showed that overall items were aligned to the standards, both content and performance. When it came to the challenge criterion, the quality of items seemed to be good with some states having their items geared towards low challenging objectives and for the balance criterion, tests in most states were not balanced well, focusing on less important standards and objectives (Resnick et al., 2004). States also scored low on the range criterion suggesting that tests did not sample a representative range of objectives intended to be measured. Identifying the issues in standards-assessments alignment in the states studied, helps states work harder and focus on specific elements to improve and enhance alignment.

Surveys of Enacted Curriculum (SEC) Model

Porter and Smithson developed The SEC alignment methodology to help educators see if what is taught in the classroom is in agreement with what is assessed, as it compares the degrees of alignment between standards, assessments, and instruction (Martone & Sireci, 2009). The SEC method assesses the alignment between the standards and assessments using a two-dimensional matrix covering content topic and cognitive demand. A group of content experts mainly do the coding of the objectives and the assessment items into the matrix, however, it is possible for teachers to do it as well (Martone & Sireci, 2009). The SEC method uses five levels

of cognitive demands: memorize, perform procedures, communicate understanding, solve non-routine problems, and conjecture/generalize/prove (Martone & Sireci, 2009). Unlike the Webb and Achieve Model, the SEC model measures instructional content using surveys that teachers fill in, and then the proportion of each topic in relation to the total instructional time is determined (Martone & Sireci, 2009). The survey approach is inexpensive and can collect information from a large number of teachers compared to other data collection approaches such as classroom observations; however the approach can face self-report bias from teachers especially because they fill in the surveys at the end of the year, hence it might not be accurate (Martone & Sireci, 2009).

Application of the Surveys of Enacted Curriculum Model. Rolf Blank and his colleagues used the SEC method to study the alignment between instruction and assessments across six states (as cited in Martone & Sireni, 2009, p. 1349). Test items were coded by at least 4 independent raters where they matched each item to a maximum of three topic areas. The added dimension in this method, the instructional content dimension, was evaluated by surveying 600 teachers from 200 schools within the six states. An alignment index was calculated and the average alignment between instruction and state assessment ranged from 0.23 in grade 8 science to 0.42 in grade 4 math. An index value ranges from 0 to 1; however, the SEC method doesn't specify a cutoff value to represent an acceptable alignment. Instead, the alignment index is viewed in relation to the other components being studied, for example comparing the alignment of state test with state curriculum versus the alignment of state test with other standards. The study was done before the NCLB movement when state assessments were not high-stake tests and therefore instruction was not highly affected compared to post-NCLB where teachers taught to the test. The indices calculated in the study do not provide enough information about the

nature of misalignment, nor does it provide more detail beyond content and depth match like the Webb and Achieve methods.

Lebanese Science Curriculum and Alignment Studies

In this study, the Lebanese science curriculum, developed in 1998, was used to investigate whether classroom assessments used in Lebanese schools are aligned to the curricular learning objectives. According to the Center for Educational Research and Development (CERD) (1995), the curriculum used before 1998 was "outdated in terms of content, lacked general and specific objectives, and was mainly focused on the theoretical rather than the practical aspects of knowledge" (cited in BouJaoude, 2002, p. 145) and therefore a new science curriculum, which is still used currently, was developed.

The main part of the fourth grade Lebanese science curriculum consists of four columns: content, learning objectives, activities & materials, and remarks. The first column, 'Content', includes five headings under which there are subheadings that provide general topics of the content. In the second column, 'Learning objectives', the competencies and skills that are expected from students to achieve are developed as objectives. In the third column, 'Activities & Materials', some experimental and documental activities are provided as suggestions for teachers to do. Additional remarks are stated in the fourth column, 'Remarks'.

The fourth grade science curriculum is divided into 5 themes: (1) Plants and their habitat; (2) Animals and their habitat; (3) Man and his health; (4) Matter and Energy; (5) Earth and the Universe. The 120 science periods are divided over these themes into 27 periods for the first theme, 21, 15, 40, and 17 periods for the second, third, fourth and fifth theme respectively.

The curriculum is set by the Ministry of Education and Higher Education; however, most of the textbooks schools follow are commercially produced to meet the curriculum specifications (Marlow-Ferguson, 2002). Public and private schools were free to choose their textbooks up until 1970s where the government adopted the books created by CERD and therefore only public schools have to use them while private schools can still choose their own textbooks in all subjects except the civics area where they are required to use the CERD's textbooks (Marlow-Ferguson, 2002).

While searching for alignment studies using the common alignment approaches, there were only the ones done in the USA and they were studies at the national level comparing national standards to national examinations (Resnick et al., 2004; Martone & Sireci, 2009; Webb, 1999; Webb, 2002). No alignment studies using common alignment approaches at classroom level were found. Furthermore, only three alignment studies in Lebanon were found **(**Hajo, 2010; Osta, 2007; Sleiman, 2012).

A Ph.D. study by Hajo (2010), investigated the degree of alignment in content and cognitive demands between the Lebanese national chemistry examinations and the chemistry curriculum at secondary level (Grade 12). The study investigated the consistency of the chemistry national exams in content and cognitive demands over a three-year period, from 2004 to 2006. Also, the Webb alignment model was employed in the study where nine experts in chemistry reviewed the curriculum and the three exams collected to judge the alignment. Data was collected from experts analyzing the documents, and by using survey questionnaires. Results have shown that the Webb alignment model is not appropriate for Lebanon, the national examinations do not measure the content of the curriculum adequately, and the examinations are not consistent in their cognitive demands from year to year.

Another study in Lebanon by Osta (2007) investigated the alignment of the pre-reform Lebanese national math examination with the pre-reform curriculum at middle school level. The study is part of a wider study that investigates the extent the older testing culture is influencing the reformed curricula. A methodological framework was developed for analyzing the pre-reform official math exams, taking into consideration the integrated and complex nature of math test items (Osta, 2007). A double-entry matrix was used where frequencies and percentages of test items addressing content areas or cognitive abilities were compared, providing results on content coverage and cognitive levels addressed in the tests. Correlational techniques were used to measure alignment between the official exams and curriculum by comparing the distribution of test items in the official exams with those in the model tests. Osta (2007) compared the national examinations to the model examinations that represent the curriculum; results indicated that there was lack of alignment. In addition, when comparing the national examinations over six years, results showed that the tests have stable structure in format and in content, and the topics covered were limited.

Another recent study by Sleiman (2012), investigated the alignment between the official Lebanese math exams and the math curriculum at secondary level. Both quantitative and qualitative analyses were used as methods to collect data, and semi-structured interviews were conducted with two test developers of the official exams. Sleiman (2012) adopted the design of analysis from Osta (2007), but used different cognitive domains for classification; she used TIMSS cognitive domains while Osta (2007) used the NAEP mathematical abilities. Results showed that there was low positive correlation between the model tests and the official exams when detailed objectives were considered; however when math topics and the cognitive domains were considered, high positive correlations were found (Sleiman, 2012). Qualitative analysis

revealed problems in coverage as both the official exams and the model tests, which reflect the curriculum, neglect some topics from the curriculum and focus on the cognitive domain 'knowing' and 'applying' while neglecting 'reasoning' (Sleiman, 2012).

All alignment studies found were on national level comparing official curriculum to national examinations at middle and secondary level in chemistry and mathematics. For this reason, this study will tackle the scarcity of information and will investigate the alignment between the official Lebanese science curriculum and classroom assessment at elementary level in schools that follow the Lebanese curriculum.

Summary

Educational assessment should be seen as an integral part of the teaching and learning process. Since assessments are the basis for important educational decisions, they need to be valid, reliable, and fair. According to the 1990 'Standards for Teacher Competence in Educational Assessment of Students' teachers need to have basic competencies regarding assessments such as being skilled in developing assessments that are appropriate for instructional decision (AFT, NCME, & NEA, 1990). One important step in constructing classroom assessments involves preparing items or tasks that are a representative sample of the domain assessed, and therefore teachers need to have the skills to align their classroom assessments to the learning objectives. Alignment is also needed for content validity, to produce accurate measures of student attainment of objectives, and to make fair decisions.

To align assessments to learning objectives within a curriculum, there are several methods mentioned in the literature; some alignment studies focused on comparing the content of the standards to the assessment (Webb, 1999; Webb, 2002) and other studies, such as the one

using the SEC model, included the instructional dimension as well. However a preferred method for determining the degree of alignment has not yet been agreed upon since different models have different focuses and each has its strengths and limitations in specific situations (Martone & Sireci, 2009). Nonetheless, all reviewed methodologies start with the basic evaluation of the alignment of the content and cognitive-level between curricular learning objectives and assessments.

When deciding which best approach or methodology to use, the SEC method was eliminated because the main goal of this research is not to study the enacted curriculum (instruction), but the assessed curriculum. When comparing the Webb and Achieve approach, the latter methodology provides more qualitative information about the overall alignment while the Webb methodology provides the most detailed quantitative results as it specifies more guidelines and cutoff scores as to what are acceptable levels of alignment (Martone & Sireci, 2009). Both methods are very similar but neither will be adopted as these common alignment approaches were designed to fit the US context and their standards system, which cannot be applied to this study because not only are the assessments at classroom level, but also the Lebanese curriculum does not adopt the standards system. The next chapter will describe the alignment method that is used in this study.

CHAPTER III

Methodology

It is important to understand the resources available when deciding which alignment approach to use and the ultimate purpose of the research or study. This study's main purpose was to investigate the alignment between classroom assessments and the Lebanese science curriculum using two criteria: content and cognitive level. Specifically, the study aimed to answer the following research questions:

1) To what extent do classroom assessments measure the content of the fourth-grade Lebanese science curriculum?
    a. To what extent do classroom assessments measure the content of the curriculum?
    b. How are the assessment items in classroom assessments distributed among the content topics in the curriculum?
2) To what extent are the cognitive levels of items in classroom assessments aligned with the cognitive levels required by the learning objectives as stated in the fourth-grade Lebanese science curriculum?

In this chapter, the design of the study is first discussed. Information about the sampling procedure, instruments that were used, phases of the study and procedures, and the data analysis are presented.

Research Design

The research design for this study is descriptive and involved the analysis of artifacts that consist of the fourth-grade Lebanese science curriculum and the summative science classroom assessments. The analysis was done by the researcher and then reviewed by two experts in the

field. The specific information that was targeted was specified ahead of time before data collection (Gall, Gall, & Borg, 2014).

The artifacts were analyzed both qualitatively and quantitatively. The curriculum and assessments were analyzed in terms of their content, objectives that are covered, and the cognitive levels they measure. Descriptive statistics were used to answer the questions raised in this study. The study consists of three phases: first the curricular learning objectives were coded according to their content and were assigned a cognitive level in consultation with the reviewers, while in the second phase, the assessments were coded and each item was matched to a content topic, a learning objective, and assigned a cognitive level, and in the third phase the alignment between the assessments and curriculum was analyzed and reported; this is further explained in the 'Procedures' section.

Sampling Procedure

All the primary data sources in this study were artifacts/documents that consist of a sample of the fourth-grade Lebanese science curriculum and the summative science classroom assessments from schools that follow the Lebanese curriculum in fourth-grade.

Theme Selection. Regarding the curriculum, there is only one official Lebanese science curriculum for the fourth grade and only one theme was used in this study. The science curriculum is divided into 5 themes: (1) Plants and their habitat; (2) Animals and their habitat; (3) Man and his health; (4) Matter and Energy; (5) Earth and the Universe. There are 120 science periods allocated to cover the whole curriculum. The 120 science periods are divided over these themes into 27 periods for the first theme, 21, 15, 40, and 17 periods for the second, third, fourth and fifth theme respectively. The fourth theme, 'Matter and Energy' was the chosen theme for

the study because it is the most comprehensive one which comprises of 35% of the curriculum's learning objectives, and constitutes an estimated 33% of the total allocated science time. There are ten topics under the 'Matter and Energy' theme which consist of definition of matter, properties of matter, measurement of mass, mixtures and water solutions, magnets, the electric charge, sound and some of its properties, propagation of sound, how do we hear, and effect of noise on our health.

Selection of Schools and Assessments. Regarding the assessments, grade 4 summative assessments regarding 'Matter and Energy' were collected from a sample of schools. The population was all of the public or private schools in Beirut that have elementary classes, follow the Lebanese science curriculum at fourth-grade and where English is the language of science instruction. To select a representative sample of the population, stratified sampling was used. Beirut is divided into three regions and the ratio of schools between regions 1, 2, and 3 is 1:7:3; this was calculated by counting the number of schools that have elementary classes and English is the language of science instruction. In addition, the ratio of public to private schools was calculated for each region. In Beirut regions 1, 2 and 3, the ratios of public to private schools are 1:5, 1:2, and 1:4 respectively (see Table 1). In this study, the researcher used stratified sampling where 18 schools were randomly selected within each stratum (Beirut region), two schools were randomly selected from Beirut region 1, eleven schools were randomly selected from Beirut region 2, and five schools were randomly selected from Beirut region 3 (see Table 1). Schools in each stratum were selected based on the ratio of public to private schools. The table below shows the distribution of schools in each region, and the number of schools that were selected from each region. The final sample, however, became 17 schools because in Beirut region 1, only one school out of the six schools agreed to participate in the study.

Table 1

Distribution of schools based on Beirut regions

Beirut Regions	Number of Public Schools	Number of Private Schools	Total number of schools	Ratio Public: Private	Distribution of schools selected		Total number of schools selected (18)
					Public	Private	
1	1	5	6	1:5	0	2	2
2	14	30	44	1:2	4	7	11
3	3	12	15	1:4	1	4	5

After selecting schools, all summative science assessments assessing the unit 'Matter and Energy' were requested from all Grade 4 classroom teachers in the selected schools for analysis. From the 17 schools, 42 assessments were collected in total. The table below highlights the distribution of the assessments collected by the researcher among schools.

Table 2

Distribution of assessments among schools

Number of Assessments collected (per school)	Number of Schools
1	5
2	3
3	6
4	2
5	1

Instruments

There are three main data collection instruments in this study; the researcher developed two of them while the third was adapted from Sleiman (2012). The first instrument, the Curriculum Coding Sheet (CCS) (Appendix A), was used in the first phase of the study to code

every learning objective within the content topics in the unit, and for assigning a cognitive level to each one. An example of the CCS is shown in Table 3. The instrument consists of four columns. The first column (Content Topic) includes the code given to each content topic, and this is based on the number that is already assigned in the curriculum. The fourth unit was used, thus every code begins with the number four. Then the second number is given based on the order of the topic arranged in the curriculum, for example, if it is the second topic then the code will be 4.2. The second column (Learning objectives) contains the codes given to each learning objective in the curriculum (under the science unit "Matter and Energy") and specifically under which topic. For example under the first topic in the fourth unit, there are four learning objectives as stated in the curriculum; the codes that were given to those learning objectives based on their sequence are 4.1.a, 4.1.b, 4.1.c, and 4.1.d. In addition, the objective was written next to its code.

The third column (Cognitive level of learning objective) contains the cognitive level that corresponds to the learning objective as categorized by the researcher and the reviewers. Bloom's original taxonomy was adopted as a framework, where the cognitive levels consist of Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. Table 4 provides a description of the cognitive levels of Bloom's taxonomy (Miller et al., 2011; Nitko & Brookhart, 2011). Bloom's original taxonomy was chosen instead of the revised taxonomy because the curriculum was published three years before the revision was made in 2001; therefore it is more appropriate to be used. In addition, Bloom's taxonomy in general is familiar to teachers and commonly used when creating classroom assessments. So for the third column, each cognitive level was noted as a number based on its level in the hierarchy; so the number "1" corresponds to Bloom Level 1 'Knowledge', number "2" corresponds to Bloom Level 2

'Comprehension', number "3" corresponds to Bloom Level 3 'Application', number "4" corresponds to Bloom Level 4 'Analysis', number "5" corresponds to Bloom Level 5 'Synthesis', and number "6" corresponds to Bloom Level 6 'Evaluation'. The fourth column (Notes) was used to note down any comments, particularly if a cognitive level cannot be assigned to a learning objective due to the vague wording used in the curriculum.

Table 3

Example of the Curriculum Coding Sheet (CCS)

Content Topic		**Learning objectives**	**Cognitive level of learning objective**	**Notes**
4.1 (Definition of matter)	4.1.a	**Recognizes** some common examples of matter: iron, water, clay, glass, etc.…	1	
	4.1.b	**Defines** matter as a body, which has a definite chemical composition.	1	
	4.1.c	**Observes** some objects and **recognizes** the kinds of matter in them	2	
	4.1.d	**Distinguishes** between matter and objects	2	

Table 4

Bloom's Cognitive levels

Bloom's Cognitive Levels	Description
1. Knowledge	Defined as the remembering of previously learned material. Requires students to retrieve relevant knowledge from long-term memory. It involves the recall of a wide range of material, from specific facts to universals theories.
2. Comprehension	Defined as the ability to grasp the meaning of material. Represents the lowest level of understanding. Requires students to make use of the material or idea being communicated without necessarily relating it to other material.
3. Application	Refers to the ability to carry out or use learned material in new and concrete situations. It involves the use of abstractions (ideas, rules of procedure, or methods) in particular and concrete situations where they have to be remembered and applied.
4. Analysis	Refers to the ability to break down material into its constituent parts so that its organizational structure may be understood. This may include determining how the parts are related to each other and to the overall structure or purpose.
5. Synthesis	Requires students to put elements together to form a coherent whole and to reorganize elements into a new structure or pattern. I might involve the production of a unique communication, a plan of operations, or a set of abstract relations.
6. Evaluation	Requires students to make qualitative and quantitative judgments based on specific criteria and standards that are either determined by the student or given to him.

The Items Coding Sheet (ICS) (Appendix B) is the second instrument that was used in the second phase of the study. It was used to match each item of the classroom assessments to a topic and assign each item to a cognitive level. An example of the coding sheet is shown in Table 5. The instrument consists of eleven columns. The first column (Assessment Item) includes the code that represents the items in the tests collected. An item is defined as the simplest form of question asked in the assessment. So each question is considered an item and if the question had

several parts, then each part is considered an item. A Roman numeral (I, II, III…) was used as a code for every school participating in the study. Then a capital letter (A, B, C…) was used to code every assessment or test collected from each school; for example, if two tests were collected from school 'II', then the codes for the tests will be II.A and II.B. Arabic numbers (1, 2, 3…) were given to the number to which the item corresponds. If a question in the assessment contained several parts, then small Roman numerals (i, ii, iii…) were used to denote that. For example, an item code of II.A.3.ii corresponds to the item in the second part of question 3, in assessment number one, collected from school number two.

The second column (Content Topic) includes the content topic that the item is matched to. The third column (Learning objective) includes the learning objective that the test item corresponds to according to the researcher's judgment; some test items might cover more than one learning objective. The same code that was assigned to the learning objectives in phase one of the study was used. If the assessment item was matched to a learning objective then in the fourth column (Content match) a 'Yes' was written, if not then a 'No' was written.

Table 5

Example of the Items Coding Sheet (ICS)

1	2	3	4	5	6	7	8	9	10	11
Assess. item	Content Topic	Learning objective (L.O)	Content Match?	Pts.	Cognitive level of L.O	Cognitive level of item	Evaluat -ion	Cog. Level Match?	Notes of researcher	Reviewers' Comments
II.A.3	4.1	4.1.a	Yes		3	2	Lower	No		

The fifth column (Points) includes the weight that was given to the item in the test. The sixth column (Cognitive level of learning objective) includes the cognitive level of the learning objective that the item targets; this was obtained from the CCS. The seventh column includes the

cognitive level of the test item based on the researcher's evaluation of cognitive level. The cognitive levels in Bloom's original taxonomy were used as a tool to classify each learning objective to a cognitive level, and then the researcher matched the cognitive levels of test items to the cognitive level of learning objectives and assessed the alignment between them both.

In the eighth column (Evaluation), the researcher compared and decided whether the rating of the cognitive level of the item is higher, equal to, or lower than the cognitive level of the learning objective. The ninth column (Cognitive level match) is used to write 'Yes' if the cognitive levels of the learning objective and of the item were the same, and 'No' if the cognitive levels of the learning objective and of the item were not the same. The tenth column (Notes of researcher) was used to note when a test item does not fit or match the content area .The eleventh column (Reviewers' Comments) is used by the reviewers to write any comments on the matching work done by the researcher.

The third instrument (Item Analysis Sheet) was adapted from Sleiman (2012) for quantitative analysis of the assessments in the third phase of the study. The first column (Topics) includes the codes given to each topic in the unit. The second column (Learning objectives) includes the codes of all the curricular learning objectives in the science unit. The third column (Cognitive levels) includes Bloom's cognitive levels and was used to see the distribution of the test items over the cognitive levels using the points assigned to them. The fourth column (Test items) was used to list the codes of the test items that were matched to each learning objective. The fifth column (Points) was used, to sum up the number of points assessing each learning objective, based on the weight that was given to the item in the test, so that for each learning objective the total points that were matched to it in the assessment were tallied. The sixth column (Percentage) was used for each content topic to calculate the percent of points allocated to it

compared to the total points of the items that were matched to the unit; for example if from a test only 17 out of 20 points were matched to the unit, and only 6 points were matched to one of the topics in the unit then the percentage calculated is 6/17x100 = 35%.

An Item Analysis Sheet (IAS) was filled for every assessment, and each test item was analyzed as to the objective(s) it was matched to and the cognitive levels to which they were assigned. The point(s) given to each test item was the same point(s) given to them in the assessment. After completing the IAS for an assessment, total points were added and then the percentages of item points that cover each cognitive level were calculated. Also, the sixth column "Percentage" in the table was used to clarify which content topics were emphasized and which were neglected in the assessment. Table 6 shows an example of an Item Analysis Sheet.

Table 6

Example of an Item Analysis Sheet (IAS)

Topics	Learning Objectives	Cognitive Levels						Test Items	Total Points	%
		1	2	3	4	5	6			
4.1	4.1.a									
	4.1.b									
	4.1.c									
4.2	4.2.a									
	4.2.b									
Total										
Percentages (%)										

Procedures

As mentioned previously, this study involves the analysis of artifacts. The documents that were used consist of the official Lebanese science curriculum and the science classroom assessments. The study consists of three phases. In the first phase of the study 'curriculum

coding', the unit of 'Matter and Energy' from the fourth-grade science curriculum was reviewed in detail. Each content topic was coded and noted down on the CCS. Then each learning objective was given a code, and afterward, based on Bloom's Taxonomy, a cognitive level was assigned to the learning objective. The researcher completed this task, as she is familiar with the subject matter at the elementary level. She holds a Bachelor's degree in elementary education with emphasis on science and mathematics. In addition, to ensure the validity of the results two other reviewers validated her work. The reviewers were experts in the science education field. To ensure the reliability of results, the researcher met with the reviewers where she identified and explained the cognitive levels that were used. Then they discussed each objective until a consensus cognitive level was determined for each learning objective.

The second phase of the study consisted of collecting fourth-grade science classroom assessments that are summative in nature from schools in the selected sample that follow the Lebanese curriculum. Prior to collecting the assessments, school principals were contacted and informed about the study and its aims. After getting their approval, the assessments were collected and coded. Each assessment item was given a code to be able to refer back to it anytime. The code was noted down in the first column of the ICS. Initially, the researcher along with the reviewers matched a sample of five to seven assessment items in terms of learning objectives and cognitive levels. After they were done analyzing each sampled test item, the reviewers and the researcher reconciled any discrepancy in the categorization and reached consensus. The researcher then continued working on the rest of the assessment items independently, with reviewers randomly checking on work being done.

In the third phase, the assessment items were compared to the curriculum to determine if they match in terms of content and cognitive level. Using the Items Coding Sheet (ICS), for each

item that has been matched to a content topic, the cognitive level of the learning objective was noted down in the sixth column. Then in the eighth column, the matching of the cognitive level was evaluated to whether it was the same, higher, or lower, and it was noted whether there was a match or not in the ninth column. After the researcher individually coded all assessment items from all schools, a stratified random sample of 16 items was selected from all the tests collected and independently matched, by the researcher and reviewers, to the learning objectives and cognitive levels. Items were stratified according to their cognitive level so that the sample selected was distributed among Bloom's different cognitive levels (four items from each level). In addition, the content topics were divided into two groups (4.1 to 4.5 and 4.6 to 4.10) where two items were selected from each group to ensure the sample was distributed among the topics of the unit. Initially, 24 items were supposed to be selected however there weren't available items under the cognitive levels 5 and 6 and therefore the number dropped to 16. The researcher and the two reviewers each did the matching process separately. After the reviewers were done analyzing each sampled item, the percentage of agreement between the researcher and reviewers was calculated to assess inter-rater reliability. If there was no agreement, items were discussed until reaching consensus on the appropriate content topic and cognitive level of that learning objective.

After comparing and matching the assessment items to the curriculum with regards to content and cognitive levels, the Items Analysis Sheet (IAS) was filled. After completing the IAS for each assessment, total points were added and then the data were transformed into percentages relative to the total number of points in each assessment to unify the basis of comparison between different assessments. The IAS helped clarify how the item points were distributed across the content topics and across the cognitive levels. It also indicated which parts of the

curriculum were assessed and which were overlooked, and which cognitive levels were emphasized.

For every assessment, an ICS and an IAS were filled in, and percentages of assessment items that matched the content and cognitive levels were calculated to describe the alignment. From the ICS, the number of points for each assessment that successfully match the curricular content was computed as a percentage relative to the total number of points, ex. 80% of the item points were aligned in terms of content. From the ICS, the number of points that successfully matched the cognitive levels of the curriculum were counted and a percentage was computed, ex. 70% of the points were aligned in terms of cognitive level. These figures were used to address the questions raised in this study.

Data Analysis

The first part of the first question in the study focused on the extent the assessment items address the same content topics as the curriculum. To answer the question, each test item was matched to a learning objective under a content topic during the second phase of the study. For each separate assessment, the percentage of item points that were matched to learning objectives in the science unit was calculated. This indicated the extent to which the assessment items were matched to the content topics in the science unit. Sixty percent and lower is considered low alignment, between 61% and 80% is considered moderate, and between 81-100% is considered high alignment.

The second part of the first question addressed the distribution of the assessment items among the content topics. For every assessment, an Items Analysis Sheet (IAS) was filled.

However to answer this question all IAS sheets were combined into a table to allow the researcher to see which content topics were given more emphasis and which were overlooked.

The second question in the study focused on the whether the test items that matched curricular learning objectives had the same cognitive level. The researcher wanted to find out whether what is elicited from students in the assessment is as demanding in terms of cognitive level as what is stated in the learning objectives. The cognitive levels of learning objectives were compared to the cognitive levels of assessment items that have been matched to those learning objectives. The items matched to a learning objective must be at the same cognitive level specified by the objective if it were to be aligned. If the cognitive levels were higher or lower, it was noted. For every assessment, the percent of items that have their cognitive level aligned with the cognitive level of the learning objectives were calculated to answer the second question in this study. Sixty percent and lower is considered low alignment, between 61% and 80% is considered moderate, and between 81-100% is considered high alignment. In addition, the distribution of items across the different cognitive levels was presented using percentages calculated in the IAS. This enabled the researcher to identify which cognitive levels in the assessments are valued and which is disregarded.

Finally, the results were aggregated for all assessments. Descriptive statistics and frequency tables were used. Details are described in the results section.

CHAPTER IV

Results

This study's main purpose was to investigate the alignment between fourth-grade summative classroom assessments and the national Lebanese science curriculum using two criteria: content and cognitive level. To accomplish this purpose three instruments were used, two of which were developed by the researcher, the Curriculum Coding Sheet (CCS) and the Item Coding Sheet (ICS), and the third one, the Item Analysis Sheet (IAS), was adapted from Sleiman (2012). This chapter presents the results of the study.

Coding of the Fourth-Grade Science Curriculum

In the first phase of the study 'curriculum coding', the unit of 'Matter and Energy' from the fourth-grade science curriculum was reviewed in detail. As reported in Table 7, each content topic was coded and noted down on the Curriculum Coding Sheet. Then each learning objective was given a code, and based on Bloom's Taxonomy a cognitive level was assigned to the learning objective. The researcher along with two reviewers completed the coding of the curriculum, and all agreed on the cognitive levels presented in Table 7 below. Fifteen percent of the learning objectives were not classifiable as they are instructional procedures rather than learning objectives that are measurable; for these specific objectives 'NA' was noted in place of the cognitive level.

Bloom's cognitive levels 5 and 6 were not assigned to any of the learning objectives in the unit, while the majority of the learning objectives were at Bloom's cognitive level 1 (39%), then level 4 (22%), then level 2 (17%), and lastly Bloom level 3 (7%). Some learning objectives

(ex. 4.4.d) consist of two objectives, and in this case, each one was given a cognitive level; however they were considered as one objective.

Table 7

The Curriculum Coding Sheet (CCS)

Content Topic		Learning objectives	Cognitive Level
4.1 (Definition of matter)	4.1.a	**Recognizes** some common examples of matter: iron, water, clay, glass, etc.…	1
	4.1.b	**Defines** matter as a body, which has a definite chemical composition.	1
	4.1.c	**Observes** some objects and **recognizes** the kinds of matter in them	NA & 2
	4.1.d	**Distinguishes** between matter and objects	2
4.2 (Properties of matter)	4.2.a	**Observes** that objects occupy space.	NA
	4.2.b	**Infers** that two objects cannot occupy the same space under the same conditions.	4
	4.2.c	**Infers** by observation that matter has mass.	4
4.3 (Measurement of mass)	4.3.a	**Measure** the mass of an object by using a balance	3
	4.3.b	**Recognizes** the standard units of mass	1
	4.3.c	**Recognizes** some modern types of balances	1
4.4 (Mixtures and water solutions)	4.4.a	**Defines** a mixture and **gives examples**	1 & 1
	4.4.b	**Explores** the kinds of mixtures: homogenous, heterogeneous	NA
	4.4.c	**Explores** some substances which dissolve in water	NA
	4.4.d	**Defines** a solution as composed of a solvent and a solute. **Distinguishes** a true solution in a practical way	1 and 2
	4.4.e	**Concludes** that the substance of a solute remains unchanged in a solution.	4
	4.4.f	**Give examples** of non-aqueous solution.	1
4.5 (Magnets)	4.5.a	**Recognizes** the common shapes of magnets and **identifies** materials, which are attracted by magnets.	1 and 2
	4.5.b	**Explores** the parts of a magnet and **infers** the law of magnetic attraction and repulsion.	NA, 4
	4.5.c	**Recognizes** the magnetic compass and **uses it** properly. **Identifies** the magnet in some toys and gadgets.	1, 3, 2
4.6 (The electric charge)	4.6.a	**States** some aspects of electric charges around us, and **produces** electric charges by friction of suitable materials.	1, 3
	4.6.b	**Explores** the attraction and repulsion of electric charges and **infers** that like charges repel each other and unlike charges attract each other.	NA, 4
4.7 (Sound and some of its properties)	4.7.a	**Infers** that sound is produced when bodies vibrate	4
	4.7.b	**Infers** that the frequency of sound determines the sharpness of sound.	4
	4.7.c	**Concludes** that the kind of vibrating body determines the quality of sound produced.	4
	4.7.d	**Recognizes** the kinds of musical instruments and **identifies** the sound-producing element in each kind.	1 and 2
4.8 (Propagation of sound)	4.8.a	**Infers** that sound travels only in matter	4
	4.8.b	**Observes** that the speed of sound is greatest in solids, decreases in liquids, and has least speed in air.	NA
	4.8.c	**Concludes** that sound travels in all directions.	4
	4.8.d	**Recognizes** some modern sound instruments.	1

Table 7 Continued

Content Topic		Learning objectives	Cognitive Level
4.9 (How do we hear)	4.9.a	**Identifies** the parts of human ear, and **recognizes** the function of each part.	1, 1
	4.9.b	**Describes** the passage of sound within the ear and the production of sound sensation.	1
4.10 (Effect of noise on our health)	4.10.a	**Recognizes** disturbing noises.	2
	4.10.b	**States** some of the effects of noise pollution, and **identifies** personal responsibility, if any.	1 and 2
	4.10.c	**Gives examples** of modern devices for reducing noise.	1

Quantitative Analysis of the Alignment

In the third phase of the study, the assessment items were compared to the curriculum to determine if they match in terms of content and cognitive level. After the researcher coded all assessment items from all schools and completed the ICS for each assessment, the number of points assigned for each assessment that successfully match the curricular content was computed as a percentage relative to the total number of points. Also, the number of points that successfully matched the cognitive levels of the curriculum was added up and a percentage was computed. The results were aggregated for all assessments (see Table 8). Descriptive statistics were used; the percentages of alignment were categorized and these were used to generate a frequency table to describe the distribution (see Table 9 and Table 10). The median was used as it is a representative of the center of the data and is not heavily influenced by outlying measurements as the mean is. In addition, the mean cannot be used because the different assessments collected are not comparable since they are different in content and length.

Table 8

Percent of Alignment For All Assessments

School Code	Assessment Code	% of Content alignment	% of Cognitive Level Alignment	% of Lower Cognitive Level	% of Higher Cognitive Level	% of NA Cognitive Level
I	A	78	40	26	26	8
II	A	90	41	19		40
	B	100	46	26	25	3
	C	68	32	68		
	D	69	73	27		
III	A	100	58	15	5	22
	B	100	20	70		10
	C	83	48	44		8
IV	A	40	56	31		13
V	A	50	30	70		
	B	95	42	11	32	16
	C	78	39	48		
	D	95	0	68	26	5
	E	78	56	44	-	-
VI	A	26	50	10	40	-
	B	42	0	-	63	37
	C	21	0	50	50	-
VII	A	90	65	22	13	
	B	67	58	24	9	9
VIII	A	60	50	17	33	
	B	100	13	7	47	33
	C	60	50	42	8	
IX	A	10	50	50		
	B	30	17	66		17
	C	55	27	-	36	36
X	A	100	30	20		50
	B	100	18	5		77
	C	20	0	100		
XI	A	26	40	60		
	B	37	100			
	C	10	50	25	25	

Table 8 Continued

School Code	Assessment Code	% of Content alignment	% of Cognitive Level Alignment	% of Lower Cognitive Level	% of Higher Cognitive Level	% of NA Cognitive Level
XI	D	31	29	29	17	25
XII	A	20	100			
	B	3	0	100		
XIII	A	90	4		15	81
XIV	A	90	11	34	47	8
XV	A	90	28	22	50	
	B	100	20			80
	C	100	60	30		10
XVI	A	65	0	69		31
	B	100	34	41	10	15
XVII	A	67	38	50	13	
	Median:	68.5	38.5	32.5	26	17

The results of the analysis of the 42 summative assessments collected showed that the median percentage of content alignment was 68.5% (see Table 8) with the values ranging from 3 to 100 percent alignment. As shown in Table 9, 41% of the assessments had content alignment equal to or lower than 60%, which indicates low alignment. Nineteen percent of the assessments were moderately aligned between 60% and 80% content alignment, while 40% of the assessments were between 80% and 100% aligned with the content of the unit indicating high alignment. There was a wide variability among schools and within schools. For example, School X has two assessments that are perfectly aligned while the third assessment that has a content alignment of 20% only and that's because it is more comprehensive and covers more topics outside the unit.

Table 9

Distribution of Assessments and Degree of Content Alignment

% of Content Alignment	For Public Schools		For Private Schools		For All Assessments	
	Freq. of assess.	% of total no. of assess.	Freq. of assess.	% of total no. of assess.	Freq. of assess.	% of total no. of assess.
1-10	0	0	3	11	1	2
11-20	0	0	2	7	4	10
21-30	0	0	4	14	4	10
31-40	1	7	2	7	3	7
41-50	1	7	1	4	2	5
51-60	0	0	3	11	3	7
61-70	2	14	3	11	5	12
71-80	3	21	0	0	3	7
81-90	2	14	4	14	6	14
91-100	5	36	6	21	11	26
Total	14	100	28	100	42	100

There was an incidental finding when comparing the alignment of assessments between public and private schools. The median percentage of content alignment in private schools was 60% while in public schools it was 80.5%. Table 9 also breaks down the distribution of assessments and degree of Content Alignment for public and private schools. As shown in Table 9, 14% of the public schools' assessments had content alignment equal to or lower than 60%, which indicates low alignment. 35% of the assessments were moderately aligned between 60% and 80% content alignment, and 50% of the assessments were between 80% and 100% aligned with the content of the unit indicating high alignment. Meanwhile in private schools 54% of the assessments had content alignment equal to or lower than 60%, which indicates low alignment. Eleven percent of the assessments were moderately aligned between 60% and 80% content alignment, and 35% of the assessments were between 80% and 100% aligned with the content of the unit indicating high alignment.

When it comes to cognitive level alignment, the results of the analysis showed that of the items that were matched in content, the median percentage of cognitive alignment of the items was 38.5% (see Table 8). As shown in Table 10, 90% of the assessments had cognitive alignment equal to or lower than 60%, which indicates low alignment. Four percent of the assessments were moderately aligned between 60% and 80% cognitive level alignment. Only 5% of the assessments showed between 80% and 100% alignment. In addition, from the assessments that were low in cognitive-level alignment, there were assessments that had higher cognitive level items and others with lower cognitive level items. The results of the analysis showed that the median percentage of items with higher cognitive level was 26% (see Table 8); 50% of the assessments had items with higher cognitive level ranging between 8 and 50%. The median percentage of items with lower cognitive level items was 32.5% (see Table 8); 86% of the assessments had items with lower cognitive level ranging between 5 and 100%.

Table 10

Distribution of Assessments and Degree of Cognitive Alignment

% of Cognitive Alignment	Frequency of assessments	% of total number of assessments
0-10	7	17
11-20	6	14
21-30	5	12
31-40	6	14
41-50	9	21
51-60	5	12
61-70	1	2
71-80	1	2
81-90	0	0
91-100	2	5
Total	42	100

Reviewers' Agreement in Coding

Three raters including the researcher independently analyzed and matched a stratified random sample of 16 items from all the tests collected, to the learning objectives and cognitive levels. After the raters were done analyzing each sampled item, the percentage of agreement between them was calculated to assess inter-rater reliability. The overall percentage agreement for the matching of learning objectives was 100% indicating perfect agreement to assigning the items to their corresponding content topics and specific learning objectives. Table 11 presents the inter-rater agreement for the assignment of cognitive level for each of the sampled items. The overall percentage agreement for the assignment of the cognitive level was 81% indicating a high agreement between the raters.

Table 11

Inter-rater Agreement for the Assignment of Cognitive Level for each of the Sampled Items

Sampled Assess. item	Rater 1	Rater 2	Rater 3	Difference Between		
				1 & 2	1 & 3	2 & 3
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	3	3	3	0	0	0
5	2	2	2	0	0	0
6	2	2	2	0	0	0
7	4	3	3	1	1	0
8	3	3	2	0	1	1
9	2	2	2	0	0	0
10	2	2	2	0	0	0
11	4	3	3	1	0	0
12	4	3	3	1	1	0
13	4	4	4	0	0	0
14	3	3	5	0	-2	-2
15	3	3	3	0	0	0
16	1	1	1	0	0	0
Total count of 0 in difference column=				13	12	14
Total Ratings=				16	16	16
Proportion Agreement=				13/16=.81	12/16=0.75	14/16=0.88
Percent Agreement=				81%	75%	88%
Overall Percentage Agreement=					81%	

Distribution of Items Among the Content Topics

To examine the distribution of items among the content topics a Items Analysis Sheet (IAS) was filled in for each assessment. Then all the IAS sheets were combined into one table, which allowed the researcher to see which content topics were given more emphasis and which were overlooked. The results are shown in Table 12 below.

Table 12

Distribution of Items for Each Assessment among Content Topics

	Percent of item points given to Content Topics										
Assess-ments	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10.	Total
I.A			3	29	26	3	3	3	15	18	100
II.A	4	5	44	47							100
II.B				23	29	34	15				101
II.C	2		18		11	16	20	11	22		100
II.D	6	10	84								100
III.A	5	5	10	80							100
III.B					50	50					100
III.C	4	20		12	4	16	40		4		100
IV.A	6	6	13	31	44						100
V.A	50	50									100
V.B	11	5	11	74							101
V.C					71	29					100
V.D							58	42			100
V.E					21		21	23	10	26	101
VI.A		10	90								100
VI.B				100							100
VI.C						100					100
VII.A	46	7	48								101
VII.B	23	3	34	23	11	6					100
VIII.A	36	6	58								100
VIII.B			7	88	5						100
VIII.C	17	13	71								101
IX.A						100					100
IX.B					100						100
IX.C				50	50						100
X.A				50	50						100
X.B				77	23						100
X.C		50	50								100
XI.A			100								100
XI.B				100							100
XI.C			100								100
XI.D	4	4	50	42							100
XII.A				100							100
XII.B		100									100
XIII.A				100							100

Table 12 Continued

Percent of item points given to Content Topics											
Assess-ments	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10.	Total
XIV.A	28	36	36								100
XV.A	83	11	6								100
XV.B				100							100
XV.C					100						100
XVI.A							31	69			100
XVI.B					100						100
XVII.A					88	13					101
Mean	8	8	20	26	18	9	4	3	1	1	98

Figure 1 illustrates the mean distribution of items among each content topic. Results of the analysis showed that the topic ‘Mixtures and Water Solutions’ (26%) gets the highest percentage out of the test items. Next is ‘Measurement of Mass’ (20%), then ‘Magnets’ follows (18%). These content topics were highly assessed compared to topics 1 (Definition of matter), 2 (Properties of Matter), and 6 (The Electric Charge) that were moderately assessed with 8-9%. Topics 7 (Sound and some of its properties), 8 (Propagation of Sound), 9 (How do we hear), and 10 (Effect of noise on our health) were the least topics to be assessed with percentages 4, 3, 1, and 1 out of the test items respectively.

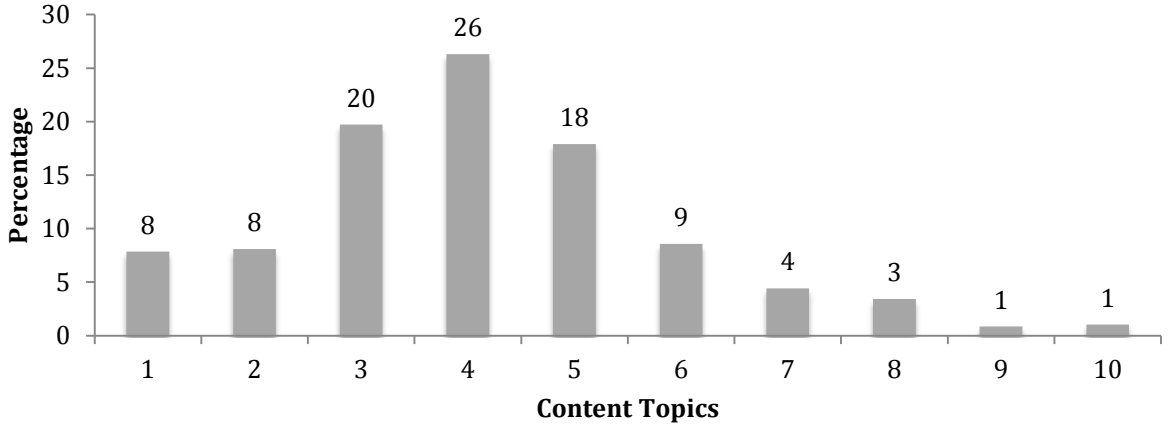

Figure 1. Mean distribution of assessment items among content topics. 1=Definition of Matter, 2= Properties of Matter, 3= Measurement of Mass, 4= Mixtures and Solution, 5= Magnets, 6= The Electric Charge, 7= Sound & properties, 8= Propagation of Sound, 9= How do we Hear, 10= Effect of noise on health.

Further analysis to see the difference in content topic distribution between public and private schools is illustrated in Figure 2 below. The public schools appear to have a wider distribution of items that cover all topics while private schools appear to have a narrower distribution of items that cover fewer topics than public schools.

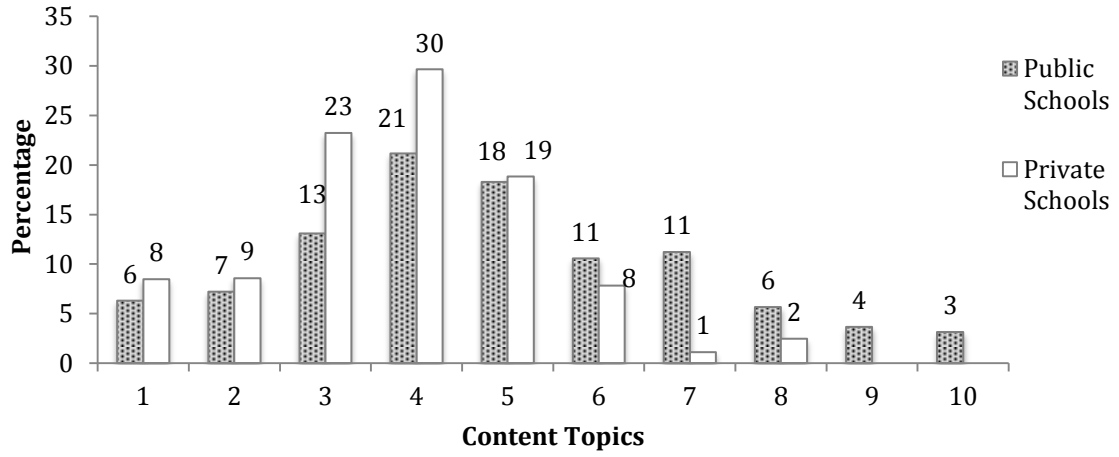

Figure 2. Mean distribution of assessment items among content topics (Public vs. Private Schools). 1=Definition of Matter, 2= Properties of Matter, 3= Measurement of Mass, 4= Mixtures and Solution, 5= Magnets, 6= The Electric Charge, 7= Sound & properties, 8= Propagation of Sound, 9= How do we Hear, 10= Effect of noise on health.

Distribution of Items Among the Cognitive Levels

The distribution of items across the different cognitive levels was described using percentages calculated in the IAS. The results were aggregated for all assessments into the table below (Table 13). The table illustrates which cognitive levels in the assessments are valued and which are disregarded.

Table 13

Distribution of Items for Each Assessment among the Cognitive Levels

	Bloom's Cognitive Levels						
Assessments	1	2	3	4	5	6	Total
I.A	60	40					100
II.A	12	55	33				100
II.B	35	50	15				100
II.C	59	41					100
II.D	42	14	43				99
III.A	35	65					100
III.B	60	40					100
III.C	40	40	20				100
IV.A	88	12					100
V.A	90	10					100
V.B	84	16					100
V.C	37	42	21				100
V.D	42	47	11				100
V.E	31	64	5				100
VI.A	40		60				100
VI.B			25	50	25		100
VI.C		100					100
VII.A	36	47	18				101
VII.B	49	32	12	6			99
VIII.A	11	17	39	33			100
VIII.B	7	75	13	5			100
VIII.C	13	46	42				101
IX.A	50	50					100
IX.B	50		33	17			100
IX.C		64	36				100
X.A		50	30	20			100
X.B	25	15	60				100
X.C		100					100
XI.A		60	40				100
XI.B	14	86					100
XI.C		25	50	25			100
XI.D	8	50	25	17			100
XII.A		100					100
XII.B	100						100
XIII.A	4	85	11				100
XIV.A		53	39	8			100
XV.A	39	61					100
XV.B	40	60					100
XV.C	50	50					100
XVI.A	23	35	34	8			100
XVI.B	30	44	10	16			100
XVII.A	25	75					100
Mean	32	46	17	5	1	0	101

Figure 3 below illustrates the mean distribution of items for each cognitive level. Results of the analysis showed that almost half of the items (46%) address the cognitive level "Comprehension", 32% address the cognitive level "Knowledge", while only 17% address the cognitive level "Application" and 5% address the cognitive level "Analysis". The cognitive levels "Synthesis" and "Evaluation" were barely addressed with percentages 1% and 0% respectively.

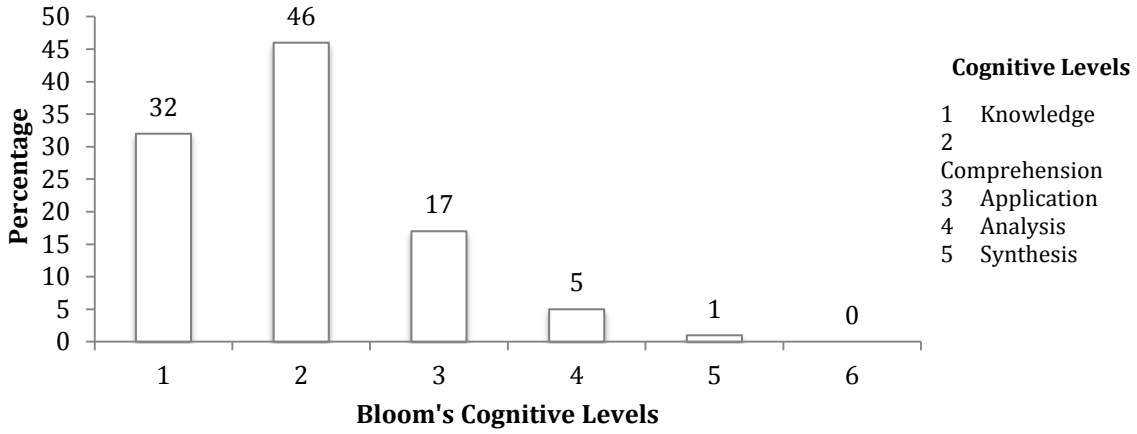

Figure 3. Mean distribution of assessment items among Bloom's cognitive levels.

Further analysis to see the difference in cognitive-level distribution between public and private schools is illustrated in Figure 4 below. The public schools concentrated on three of Bloom's cognitive levels which were: 'Knowledge' (51%), 'Comprehension' (38%), and then 'Application' (11%). On the other hand the private schools had a wider distribution across the cognitive levels; they concentrated mainly on 'Comprehension' (49%), then 'Knowledge' (22%) and 'Application' (21%), then 'Analysis' (7%), then 'Synthesis' (1%). Neither type of schools tackled Bloom's cognitive level 'Evaluation'.

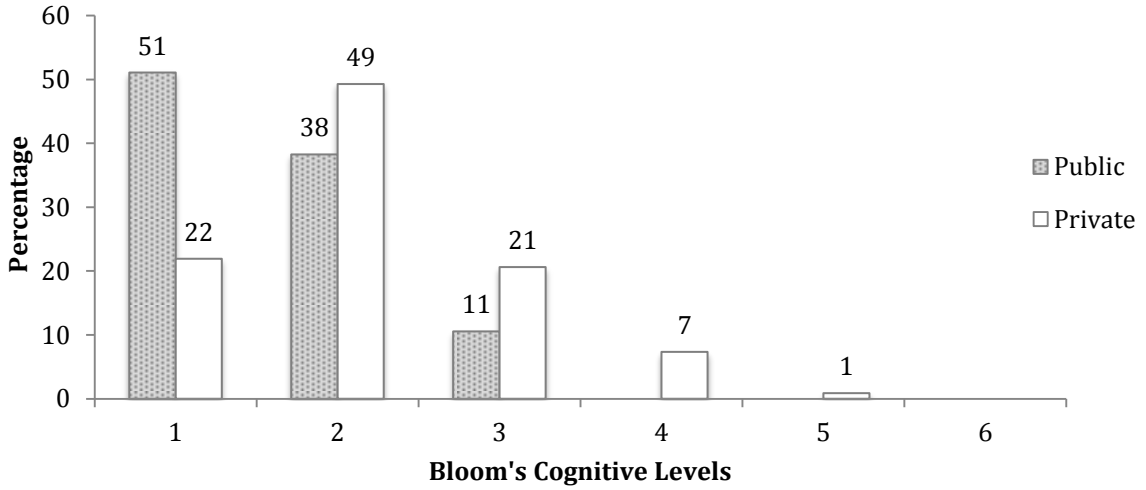

Figure 4. Composition of the mean distribution of assessment items among Bloom's cognitive levels between public and private schools. 1=Knowledge, 2= Comprehension, 3= Application, 4= Analysis, 5= Synthesis, 6=Evaluation.

Furthermore, the difference between the cognitive levels assigned to the curricular learning objectives and the cognitive levels assigned to the assessment items was analyzed and the results are illustrated in Figure 5 below. The curriculum concentrated on Bloom's cognitive levels of 'Knowledge' (39%), 'Analysis' (22%), and 'Comprehension' (17%) while the assessment items concentrated on Bloom's cognitive levels of 'Comprehension' (46%) and 'Knowledge' (32%), and 'Application' (17%). Cognitive levels "Synthesis" and "Evaluation" were barely addressed in both the Curriculum and assessments.

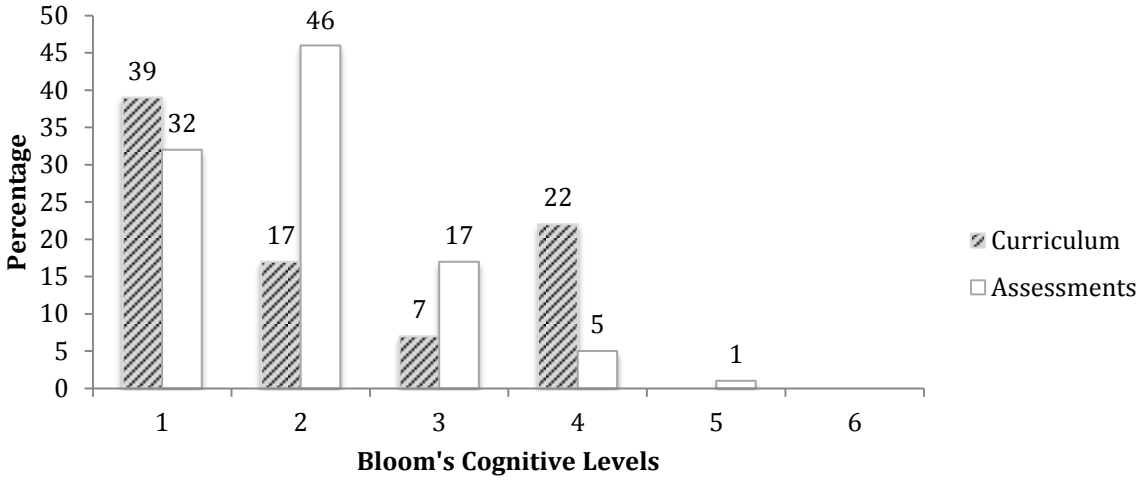

Figure 5. Cognitive Level Assignment (Curriculum vs. Assessments). 1=Knowledge, 2= Comprehension, 3= Application, 4= Analysis, 5= Synthesis, 6=Evaluation.

Summary of Results

The data collected from the three instruments (CCS, ICS, and IAS) show that when coding the curriculum, no objectives were assigned to Bloom's cognitive levels 5 and 6, while majority of the learning objectives were assigned to Bloom's cognitive level 1 (39%), then level 4 (22%), then level 2 (17%), and lastly Bloom level 3 (7%); the rest (15%) were not applicable.

The results of the analysis of the 42 summative assessments collected showed that the median percentage of content alignment was 68.5%, while the median percentage of cognitive-level alignment of the items was 38.5%. From the assessments that were low in cognitive-level alignment, there were assessments that had higher cognitive level items and others with lower cognitive level items. The median percentage of items with higher cognitive level was 26%, while the median percentage of items with lower cognitive level items was 32.5%. Thus, 65% of items are at the same level of cognitive-level or higher, which indicates a moderate level of agreement.

When analyzing the distribution of items across content topics, results showed that the topics ‘Mixtures and Water Solutions’, ‘Measurement of Mass’, and ‘Magnets’ were mostly assessed while Topics ‘Sound and some of its properties’, ‘Propagation of Sound’, ‘How do we hear’ and ‘Effect of noise on our health’ were the least to be assessed. Further analysis showed that the public schools appear to have a wider distribution of items that cover all topics while private schools appear to have a narrower distribution of items that cover fewer topics than public schools (Figure 2).

In addition, when analyzing the distribution of items across Bloom’s cognitive levels, results showed that almost half of the items address the cognitive level “Comprehension”, then “Knowledge”, then “Application”. The cognitive levels “Analysis” and “Synthesis” where least addressed whereas “Evaluation” was not addressed at all. Private schools covered more cognitive levels than the public schools, which concentrated on only 3 cognitive levels: Knowledge, Comprehension, and Application. Half of the items from public schools addressed ‘Knowledge’ while half of the items from private schools addressed ‘Comprehension’. Furthermore when comparing the cognitive level distribution between the curriculum and the assessments, results showed that the curriculum concentrated on Bloom’s cognitive levels of ‘Knowledge’, ‘Analysis’, and ‘Comprehension’, while the school assessment items concentrated on Bloom’s cognitive levels of ‘Comprehension’, ‘Knowledge’, and ‘Application’. The cognitive level “Synthesis” was least addressed (only by assessments), whereas “Evaluation” was not addressed at all by either the curriculum or assessments.

The next chapter presents the discussion of findings from the analyzed data and answers to the different research questions.

CHAPTER V

Discussion

Classroom tests and assessments play a crucial role in evaluating student learning and achievement. They are also the basis for important educational decisions thus they need to be valid, and reliable. A basic requirement for the validity of classroom assessment results is that the assessment measured should be aligned with the curriculum's learning objectives (Nitko & Brookhart, 2011). In addition, assessments should match the scope of content covered in the learning objectives, and should match the depth of thinking and cognitive skills required by the objectives (Nitko & Brookhart, 2011).

While searching for alignment studies using the common alignment approaches, there were only studies conducted in the USA and they were at national level comparing national standards to national examinations (Martone & Sireci, 2009; Resnick et al., 2004; Webb, 1999; Webb, 2002). No alignment studies using common alignment approaches at classroom level were found. Furthermore, only three alignment studies in Lebanon were found **(**Hajo, 2010; Osta, 2007; Sleiman, 2012), and these were on national level comparing official curriculum to national examinations at the middle and secondary level in chemistry and mathematics. For this reason, this study's main purpose was to investigate the alignment between the summative classroom assessments at the elementary level in schools that follow the Lebanese curriculum and the official Lebanese fourth-grade science curriculum. Specifically, the study aimed to answer the following research questions:

1) To what extent do classroom assessments measure the content of the fourth-grade Lebanese science curriculum?
    a. To what extent do classroom assessments measure the content of the curriculum?

b. How are the assessment items in classroom assessments distributed among the content topics in the curriculum?

2) To what extent are the cognitive levels of items in classroom assessments aligned with the cognitive levels required by the learning objectives as stated in the fourth-grade Lebanese science curriculum?

Research Question 1 (First Part)

The assessment items were compared to the curriculum to determine if they match in terms of content since assessments should match the scope of content covered in the learning objectives (Nitko & Brookhart, 2011). Overall the median percentage of content alignment of 68.5% indicates moderate alignment. There was an incidental finding when comparing the alignment of assessments between public and private schools. The median percentage of content alignment in private schools was 60% while in public schools was 80.5%. It is important to mention that the assessments collected by the researcher weren't all the same in terms of length; in other words, some assessments covered more topics than others. Some schools handed in one end-of-semester exam (example assessment from school I); while other schools do not follow such systems but rather do one exam at the end of each unit they cover (example assessment from school XVI); thus these will be more aligned than the other exams that cover other topics taken throughout the semester (example assessments from school III). This explains the wide range of values the researcher recorded for content alignment percentages.

With respect to the incidental finding in relation to public/private significant differences in alignment, it was not surprising to find out that only 14% of public school assessments had content alignment equal to or lower than 60%, and 50% of the assessments were highly aligned (between 80-100%), while in private schools 54% of assessments had content alignment equal to

or lower than 60%, and only 36% were highly aligned, since public schools follow the national curriculum without referring to other sources, and in addition they use the textbooks published by CERD, which follows the national Lebanese curriculum (Marlow-Ferguson, 2002; Osta, 2007).

Research Question 1 (Second Part)

After completing the IAS for each assessment and combining the data, it was easy to see the distribution of the assessment items in classroom assessments across the content topics in the curriculum (Figure 1). Findings showed that the topics Mixtures and Water Solutions, Measurement of Mass, and Magnets, were awarded the highest percentage out of the test items relatively to others while least topics to be covered were Sound and some of its properties, Propagation of Sound, How do we hear, and Effect of noise on our health. When analyzing the distribution of learning objectives in the curriculum among the content topics, there was no alignment or relationship between the two distributions; for example topics 1, 7, and 8 had equal number of learning objectives (12%) that cover 36% of the unit, however in the assessments, only 15% of items covered these topics and the items were not distributed equally among them. The only exception was the topic Mixtures and Water Solutions, as it was highly assessed and also had highest number of learning objectives (18% of objectives) compared to the rest. Furthermore, the number of curricular learning objectives under each topic might not imply which has more content; for example the topic Mixtures and Water Solutions is wider and covers more material than the topic How Do We Hear. For this reason, it would be useful to also analyze the amount of material in the books to see which topics cover more material.

When comparing the difference in content topic distribution between public and private schools, the public schools appear to have a wider distribution of items that cover all topics while

private schools appear to have a narrower distribution of items that cover fewer topics than public schools; hence this shows how the public schools are more aligned with the Lebanese curriculum than private schools are. According to Osta (2007):

> *In Lebanon, a national curriculum is in effect, which is binding to both public and private schools. While public schools implement only the national curriculum and textbooks, private schools may implement more than one program and may use different series of textbooks but are bound to teach also the national curriculum. (p. 175)*

This was shown in the study as some private schools used the national curriculum as well as other foreign curricula to follow, thus they chose what topics to cover and what to overlook. For that reason, there were some topics (topics 7, 8, 9, and 10) that were covered in the curriculum but were not addressed in the assessments provided.

Even in public schools at this level, teachers are capable of skipping topics or not giving them much importance especially if they are running out of instructional time. The elementary level is not considered a serious one where all objectives should be met such as in grades nine and twelve where national examinations take place; even at those grades (9 and 12), some topics are valued more than others. For example in Sleiman's study (2012), results revealed that both the secondary-level mathematics official exams and the model tests neglect some topics from the curriculum. Also, results from Resnick and his colleagues' study (2004) showed that tests in most states were not balanced well, focusing on less important standards and objectives. However, in this study, it is difficult to conclude which content topics are more important than others. In Osta's study (2007), findings showed that the Lebanese national examinations at ninth grade level in math were stable in terms of content coverage and abilities addressed and that they had a low level of mobility from one test to another; this stability makes the topics covered in

class shrink gradually focusing only on the content that is valued by the exams. However, since there are no national examinations at the elementary level, there aren't external forces that guide teachers to focus on some topics more than others; this is shown in this study by the wide variety of content topic coverage from one school to another and between public and private schools.

An important point is that even though topics 7, 8, 9, and 10 were barely covered by private schools, yet this does not mean that they are not covered at all; some schools teach these topics at other grade levels. It is important to mention that some schools take the Lebanese curriculum and they distribute the learning objectives based on their students' level; they teach a specific topic at the grade that suits their students, so a topic could be taught at grade 4 in one school while it could be taught at grade 3 in another school. When the researcher was collecting science assessments from schools, many schools said that they do not cover some curricular topics at the fourth-grade level. When they were asked informally about the reason why they do not cover these topics, their reply was that their aim is to fulfill the objectives at the end of the cycle rather than the grade. Even when analyzing the assessments from private schools, there were topics such as properties of matter, volume, and forces that were addressed at fourth-grade but in the national curriculum, these were supposed to be covered in third grade. Also the topic 'physical and chemical change' was covered in a fourth-grade assessment while it was supposed to be covered at sixth grade according to the national curriculum.

Research Question 2

The assessment items were compared to the curriculum to determine if they match in terms of cognitive level since assessments should match the depth of thinking and cognitive skills required by the objectives (Nitko & Brookhart, 2011). For each assessment the number of points that successfully matched the cognitive levels of the curriculum were counted and a

percentage was computed. After the results were aggregated and analyzed for all 42 summative assessments, findings showed that the median percentage of cognitive-level alignment was 38.5%. The median percentage of items with higher cognitive level was 26%, while the median percentage of items with lower cognitive level items was 32.5%. A mean percentage of 38.5% is considered low, however, there were assessments that were low in cognitive-level alignment but they had higher cognitive-level items, and others with lower cognitive-level items. If we roughly add the percentages of items that were aligned to the ones with higher cognitive-level we would get around 65% of items that are at the same level of cognitive-level or higher which indicates a moderate level of agreement. This percentage agrees with Hajo (2008) who states if the curriculum emphasizes low cognitive-level thinking then there is a high probability that the cognitive-level of the assessment items can be at or above the cognitive level of the objectives they were matched to. It is important to mention that a median percentage of 17% was not classified as same, higher, or lower because a cognitive level was not assigned to the learning objective the item was compared to.

Further analysis was done regarding the cognitive levels of the items. The researcher analyzed the distribution of items across Bloom's cognitive levels. Results showed that most of the items addressed the cognitive level "Comprehension" and "Knowledge", while cognitive levels "Synthesis" and "Evaluation" were barely addressed. Similarly, in Sleiman's study (2012) results revealed that both the official exams and the model tests at secondary-level focused on the cognitive domain 'knowing' and 'applying' while neglecting 'reasoning'. Also in Webb's study (1999) the assessment items generally targeted low cognitive-level.

In addition, the researcher analyzed the difference between the cognitive levels assigned to the curricular learning objectives and the cognitive levels assigned to the assessment items

(Figure 5). Even though relatively the curriculum had a high percentage for 'Analysis' (22%), but only 5% of the assessment items were on that level, and these assessments came from private schools only. One of the explanations could be that when these learning objectives are taught in class they become low cognitive demanding. For example the learning objective, "infers that like charges repel each other and unlike charges attract each other" was assigned to Bloom's level "Analyzing"; however during instruction once students infer that through an experiment or demonstration, it becomes a fact that they know; hence if they were asked about it in an assessment (for example a fill in the blank question: Like charges ____ and unlike charges____) then it might become Bloom level one or two (depending on the question) as they are not inferring anymore since they've done it and now they are recalling it. The inferring cognitive procedure, which is assumed in the curriculum is done while learning and not always while assessing. This was shown abundantly in the assessments where the questions were based on recall and the learning objective they were assigned to was based on analyzing and inferring.

Findings showed that both the curriculum and the assessments devalue Bloom's cognitive levels of 'Synthesis' and 'Evaluation'. It is not sure if the assessments are to be blamed for focusing on low-level cognition since the curriculum, which they are supposed to follow, does not emphasize their use. This finding confirms the results obtained by Marlow-Ferguson (2002) who states "memorization of facts and events is greatly emphasized in Lebanese schools" (p.784). Also, this finding is somewhat similar to the results Sleiman (2012) obtained which concluded that the official exams focus on low-level cognitive domains while neglecting the higher levels.

In addition, according to the National Research Council (NRC) (2000) even though understanding has more value in science than knowledge, "it is a mistake to think that all

instruction or assessments should aim for the higher level of outcome," (p.78) and the reason is if a student does not succeed at a complex task, it would be hard to tell whether the failure was due to lack of specific skills or lack of knowledge needed unless these were examined (NRC, 2000). Therefore it is recommended to have a certain extent of low-level cognitive demand, but at the same time also to value the higher cognitive levels in curriculum design and this is the case in the Lebanese curriculum. Furthermore, having higher cognitive-level items will only be positive if the assessments are aligned to the curriculum and with classroom instruction and teaching. So even if the curriculum focuses on low cognitive levels, if the classroom instruction targeted higher cognitive thinking then assessment items with high cognitive level will be acceptable, otherwise it would be a disadvantage to students.

Assumptions and Limitations of the study

In this study, it is assumed that the sample of assessments collected is representative of all fourth-grade science summative assessments in Beirut.

The study conducted certainly has limitations. This study is specific to elementary level the results may not be generalizable to other grade levels. In addition, the results of this study are limited to the 'Matter and Energy' unit and cannot be generalized to the entire fourth-grade science curriculum. Furthermore, the sample in the study did not include schools with French as the language of instruction, only English was included. Another limitation is that some schools handed in several assessments while other schools handed in only one assessment; this may or may not reflect practice.

Conclusion

There have been few attempts in Lebanon to study the alignment between the national assessments and curriculum; however, there haven't been studies on the alignment between the national curriculum and classroom assessments at the elementary level. This study addressed this gap in research and the results showed moderate alignment between the assessments collected and the Lebanese curriculum in terms of content and cognitive level.

Content alignment was not as high, especially among private schools and they are the majority, as they do not fully follow the Lebanese curriculum as it is. Although they are bound to teach the national curriculum but they follow more than one program and might use several different series of textbooks (Osta, 2007) hence, private schools are selective in the topics they want to teach. This was demonstrated by the higher content alignment of public schools (80.5%) versus private schools (60%) because they follow the Lebanese curriculum as it is, and it is also demonstrated by the narrower content topics covered by the private schools versus public schools (refer to Figure 2).

In terms of cognitive level, there was a low alignment (38.5%) between the assessments collected in the study and the curriculum; however, there were items (26%) that were higher in cognitive-level compared to the cognitive-level of the learning objective they were matched to. In addition, 17% was not classified as 'same', 'lower', or 'higher' because a cognitive level was not assigned to the learning objective the item was compared to. Additional analysis showed that the curriculum focused on Bloom's cognitive levels of 'Knowledge' and 'Analysis' while the assessments focused on Bloom's cognitive levels of 'Knowledge' and 'Comprehension'. Even though relatively the curriculum had a high percentage for 'Analysis', conversely only a few items from private school assessments were assigned to this cognitive level. Public schools

mainly assessed at ‘Knowledge’ level and only assessed the first three cognitive levels of Bloom’s taxonomy while the private schools also included ‘Analysis’ and ‘Synthesis’ though at a low percentage.

According to Fulmer (2011), the items used in tests represent only a sample of the content domain and cognitive-level specified in the curriculum and therefore one cannot expect perfect alignment. In addition discrepancies between what the curriculum and assessments emphasize may be acceptable if the test encourages higher-order thinking than what the objectives describe (Fulmer, 2011); however, it will only be acceptable if the assessments are aligned with classroom instruction and teaching otherwise it would be a disadvantage to students. Furthermore, Fulmer (2011) states that if researchers and policymakers determine that a set of standards does not represent student outcomes adequately then a low alignment can possibly be acceptable and sometimes preferable. Therefore it is important for the curriculum to adequately present student learning outcomes in terms of content and cognitive-level for better alignment or else it would be invalid and unfair to both students and schools to judge their achievement of expectations based on a poorly aligned system of assessments (La Marca, 2001).

The Lebanese curriculum could be written in a way to make it easier for teachers to base their assessments on it; it is recommended that some learning objectives be rewritten as an end product rather than an instructional product such as the learning objective “Observes that objects occupy space” or “explores the kinds of mixtures: homogenous, heterogeneous”; the latter could be rewritten into “distinguish between homogeneous and heterogeneous mixtures”. In addition, in the curriculum there are learning objectives, such as “infers that sound travels only in matter” and “infer the law of magnetic attraction and repulsion”, that were assigned to Bloom’s level “Analyzing”; however during instruction once students infer that then it becomes a fact that they

know, hence if they were asked about it in the assessment then it becomes Bloom level one or two (depending on the question) as they are not inferring anymore. The inferring cognitive procedure is done while learning, and not while assessing. As mentioned previously, learning objectives need to be written as end products of learning instead of instructional objectives.

In addition, it is recommended that more learning objectives to be written to expand the objectives under a topic since there is an adequate amount of material to just be summed up into 2-3 learning objectives that might not even be measurable. There was material under content topics that were important and frequently assessed but there wasn't a direct learning objective that matched them. For constructing classroom tests and assessments, the first step is determining the purpose of the measure, then developing specifications, then selecting appropriate assessment tasks, and finally preparing relevant assessment tasks (Miller et al., 2011). Therefore it is recommended to expand the number of learning objectives in order to help teachers in preparing relevant assessment items that are aligned to the curriculum.

Recommendations for Further Research

The analysis that was done in this study focused on fourth-grade summative assessments and fourth-grade 'Matter and Energy' unit in the Lebanese curriculum. It is recommended that further research can be conducted to examine the assessments addressing the entire science curriculum and incorporating the whole cycle, which includes grades 4, 5, and 6. This way the researcher will be able to tell whether the school covered a specific topic or not, as sometimes even if a topic was assigned to a certain grade, the school might prefer to teach it at another grade level.

Future research may also be conducted where the analysis on a smaller sample of schools could be extended not only to include assessments, but also books and materials used for instruction and for creating the assessments. This way the researcher can assign the actual cognitive level of the items. For this study, only the CERD book could have been used; it was not possible to collect the different books and materials all 17 schools used because private schools use a wide variety of textbooks (Marlow-Ferguson, 2002).

REFERENCES

Align. (2015). *In Merriam-Webster's dictionary* (11th ed.). Springfield, MA: Merriam-Webster.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students.* Washington, DC: National Council on Measurement in Education.

BouJaoude, S. (2002). Balance of scientific literacy themes in science curricula: The case of Lebanon. *International Journal of Science Education, 24*(2), 139-156. DOI: 10.1080/09500690110066494

Bhola, D., Impara, J., & Buckendahl, C. (2003). Aligning tests with states' content standards: methods and issues. *Educational Measurement: Issues and Practice*, 22(3), 21-29.

Case, B., Jorgensen, M., & Zucker, S. (2004). Alignment in educational assessment: Assessment report. *Pearson.* Retrieved from http://www.pearsonassessments.com/NR/rdonlyres/C19A3D92-B124-4098- BC2F-0BCD490DCAA2/0/AlignEdAss.pdf

Fulmer, G. W. (2011). Estimating critical values for strength of alignment among curriculum, assessments, and instruction. *Journal of Educational and Behavioral Statistics, 36*(3), 381-402. DOI: 10.3102/1076998610381397

Gall, J., Gall, M., & Borg, W. (2014). *Applying educational research: How to read, do, and use research to solve problems of practice* (6th ed.). Essex, England: Pearson.

Hajo, Z. (2010). *Content validity and comparability of the Lebanese national examinations in chemistry: An attempt to use the Webb alignment model in Lebanon* (Unpublished doctoral dissertation). University of Leicester, United Kingdom.

La Marca, P. (2001). Alignment of standards and assessments as an accountability criterion. *Practical Assessment, Research, and Evaluation, 7*(20), 1-6.

Marlow-Ferguson, R. (2002). *World education encyclopedia: A survey of educational systems worldwide (2nd Ed).* Farmington Hills, MI: Gale Group.

Martone, A., & Sireci, S. (2009). Evaluating alignment between curriculum assessment, and instruction. *Review of Educational Research, 79*(4), 1332-1361. DOI: 10.1016/j.im.2004.02.005

Miller, M.D, Linn, R.L, & Gronlund, N.E. (2011). *Measurement and assessment in teaching (11th ed.).* New Jersey: Prentice-Hall, Inc.

Nitko, A. & Brookhart, S. (2011). *Educational assessment of students (6th ed.).* New Jersey: Prentice Hall Inc.

NRC (National research council). (2000). *Inquiry and the national science education standards: a guide for teaching and learning.* Washington, DC: National Academy Press.

Osta, I. (2007). Developing and piloting a framework for studying the alignment of mathematics examinations with the curriculum: The case of Lebanon. *Educational Research and Evaluation, 13*(2), *171-198.* DOI: 10.1080/13803610701452607

Reynolds, C, Livingston, R. & Wilson, V. (2009). *Measurement and assessment in education.* Upper Saddle River, New Jersey: Pearson.

Resnick, L., Rothman, R. Slattery, J., & Vranek, L. (2004). Benchmarking and alignment of standards and testing. *Educational Assessment, 9*(2), 1-27.

Roach, A., Niebling, B., Kurz, A. (2008). Evaluating the alignment among curriculum, instruction, and assessments: Implications and applications for research and practice. *Psychology in the Schools, 45*(2), 158-176. DOI: 10.1002/pits.20282

Sleiman, L. (2012). *A study of the alignment between the Lebanese secondary-level national math exams for the literature and humanities track and the reformed math curriculum* (Masters Thesis). Retrieved from http://search.shamaa.org/FullRecord.aspx?ID=66117

Webb, N. L. (1997). Determining alignment of expectations and assessments in mathematics and science education. *NISE Brief 1(2),* 1–8. Retrieved from http://facstaff.wcer.wisc.edu/normw/1997alignmentbrief.htm

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessment in four states*. Research monograph No.8. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (April, 2002). *An analysis of the alignment between mathematics standards and assessments for three states*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Appendix A

Curriculum Coding Sheet (CCS)

Content Topic	Learning objectives		Cognitive level of learning objective	Notes

Appendix B

Items Coding Sheet (ICS)

Assess. item	Content Topic	Learning objective (L.O)	Content Match?	Pts.	Cognitive level of L.O	Cognitive level of item	Evaluat -ion	Cog. Level Match?	Notes of researcher	Reviewers' Comments

Appendix C

Items Analysis Sheet (IAS)

Content Topics	Learning Objectives	Cognitive Levels						Test Items	Total Points	%
		1	2	3	4	5	6			
Total										
Percentages (%)										