



The Case for and Against Standardized Testing

Rabih Edriss

American University in Dubai, ridriss@gmail.com

Matthew J. Etchells

Texas A&M University, matthewetchells79@tamu.edu

Abstract

This research explores the global use of standardized testing with a focus on the United Arab Emirates (UAE) experience in utilizing The Trends in International Mathematics and Science Study (TIMSS). Five semi-structured interviews were conducted to investigate teacher perspectives regarding the use of standardized tests, specifically TIMSS. Four out of five interviewees felt preparing their students for TIMSS in 2014-2015 was a challenging experience, due to mismatch with their curricula. To cover this gap, they attempted to make their students absorb as much pertinent information as possible in a short duration which may jeopardize their ability to foster student creativity in class. It is possible aligning curricula of different schools to TIMSS topics would enhance fairness and effectiveness of academic evaluation of the schools, teachers', and students' performance. This research recommends supporting teachers and providing them with appropriate professional development and well outlined curriculum, instead of solely holding them responsible for poor student achievement on TIMSS.

Keywords: Student Achievement, Standardized Testing, Qualitative Research



Introduction

Assessment is an important tool in evaluating the outcomes of pedagogy and curriculum in education. Standardized test can be defined as any test administered, marked, and analyzed in a standard, predetermined manner (Popham, 2005). Some standardized tests such as the Trends in International Mathematics and Science Study (TIMSS), Programme for International Student Assessment (PISA), and Progress in International Reading Literacy Study (PIRLS) compare scores of students at various ages in various subjects across myriad countries. Such international assessments of student achievement are becoming increasingly common among governments around the world, as they attempt to measure the attainment of their education systems. This is because they realize the education of their citizens contributes to the economic security of their future (Baird, et al., 2011). Others evaluate schools and teachers' performance in a certain region like high-stakes testing such as in the United States of America (U.S.A.). In the last decade there has been a heated debate regarding the pros and cons of such testing. Some scholars claim the pressures of these tests stifle creativity in the classroom as teachers become more focused on ensuring all the required material for the exam is covered in time (Chomsky & Robichaud, 2014; Darling-Hammond & Turnipseed, 2015). It is forcing teachers to spend more time on test preparation and drill questions rather than genuine teaching and learning (Baker, et al., 2010). Others believe educational authorities need to assess policy achievements through standardized testing in order to improve curriculum and outcomes (Bishop, 2001; Phelps, 2005; Walberg, 2011).

Research Questions

This Study is guided by the following research questions:

1. How is the world focused on standardized testing today?
2. What, if anything, can we learn from that experience?
3. What is the United Arab Emirates perspective on standardized tests?

Literature Review

This literature review addresses the following domains: the debate on standardized testing, international standardized testing, and the UAE perspective on standardized tests.



Using Standardized Test Scores to Evaluate Teachers' Effectiveness

Using standardized test scores to evaluate teachers is a global issue as results are sometimes being used to criticize teachers performance (Kuehn, 2010) and some schools as a result were even closed down (Darling-Hammond & Turnipseed, 2015). However, there is a broad agreement among statisticians that test scores alone are not a sufficient measure to evaluate teachers' or curricula's effectiveness (Baker, et al., 2010) . Moreover, just because students are trained to do well on a test does not mean they have mastered certain skills (Chomsky & Robichaud, 2014). For example, if teachers know they will be evaluated by their students' scores on a test that predictably asks questions about triangles and rectangles, teachers will not devote much time to polygons, an equally important, but somewhat more difficult topic in the math curriculum. The rise in the scores of mathematics exams may expose, in part, teachers' improved skills in predicting the topics and types of questions, addressed by the exam instead of improved students' skills. This practice is commonly called "teaching to the test." It is a rational response to incentives like salary hike (Baker, et al., 2010). In other words, teachers who train their students to perform well in an exam get rewarded and the ones who fail to do so face sanctions. There are many factors which influence test scores, such as private tutoring, previous teachers, and level of parental support rather than only the teachers' effectiveness.

Impact of Standardized Tests on the Learning Experience of Students

Emphasis on test scores has affected the learning experience of students. For example Kohn (2000) claims USA education policy continue to stress passive learning through memorization and acknowledgment of answers on tests that accept only one right answer. Because state tests are increasingly linked to high-stakes decisions (student advancement, teacher and principal retention, school closures and funding), test preparation demands more and more classroom time. High scores would represent excessive drilling to excel in a particular test rather than learning (Kohn, 2000). The format of such tests can be considered disconnected from the real world and Kohn (2000) argues very few jobs require a right answer from memory on the spot. Additionally, collaboration skills are highly favored by employers, but would be considered as cheating during examinations. Such tests emphasize lower-level skills and drive instruction in



that direction (Darling-Hammond & Turnipseed, 2015). Furthermore, Darling-Hammond and Turnipseed (2015) concur we cannot depend on a year-end test especially one that narrows the measure of educational attainment to a single score and measures memorization and recognition. Many teachers protested against such tests (Shaeffer, 2012). The president of the American Federation of Teachers (AFT) Randi Weingarten stated it is time to restore balance in schools so the focus of education is teaching and learning rather than testing and teaching to the test (Shaeffer, 2012). Some teachers argue excessive emphasis on testing did not yield significant impact on learners. In New York State more than 1400 school directors signed a letter objecting to teachers' evaluation based on test scores and reminded the authorities that a 2011 report by the National Research stated the past decade of emphasis on testing had shown little improvement in learning (Shaeffer, 2012). Many teachers' groups and unions have suggested assessing students on real learning tasks is more accurate (Shaeffer, 2012). Popham (2014) affirms standardized tests do not improve instructional strategies as they do not accurately inform educators about the gaps in student learning.

According to Chomsky and Robichaud (2014) standardized practices like "teaching to the test" are going to destroy the American educational system rather than improving it. They also stated standardized educational practices are sort of attack on humanistic and critical education, as they are made to annihilate students and teachers creativity (Chomsky & Robichaud, 2014). The one size fits all approach is totally against the differentiation approach. They suggest it does not really matter what a child learns, because he is capable of learning what matters to him (Chomsky & Robichaud, 2014). There are questions about how suitable such exams are for students who are dyslexic or have other special needs.

Teachers perceive the current high-stakes testing climate to be negatively affecting their ability to foster creative thinking in the classroom through its emphasis on compliance and conformity at the expense of teacher autonomy and self-direction. It limits the time teachers can dedicate to attend to the various needs of their students, whereas the tests themselves favor students with certain learning styles while disadvantaging others (Olivant, 2015). Some teachers perceive high-stakes testing to disconnect them from teaching and students from learning. The time pressures imposed by tests and pacing calendars cause disconnections by eliminating the



time available for other subjects and restraining teachers' ability to respond to individual student needs. They also deprive the teachers of their sense of autonomy, thus diminishing their skill to foster creativity, by focusing on scripted curricula and regimented teaching strategies designed to improve test scores (Olivant, 2015). Chomsky and Robichaud (2014) purport because of the economic impact such tests have on the lives and choices of teachers; their salaries depending on the results of their students on standardized tests. We can conclude these policies are designed to enforce obedience, discipline, and eradication of individual initiatives, while education should depend on intrinsic motivation and personal interests of a child.

Other researchers, such as Bishop (2001), juxtapose a negative framework and claim standardized tests have made teachers work a lot harder, for example, they began tutoring during their free period and are staying late to help students. In France for example, the educational authorities responded to results of international assessments by introducing reforms of the system including personalized learning assistance for struggling students (Baird, et al., 2011). Bishop (2001) states standardized tests make struggling students get the required attention in order to make sure all students pass the exam and eventually raise the scores. Pressures of high-stakes exams are healthy as it expands peer tutoring, encourages funding for professional development of teachers and urges schools to hire teaching assistants (Bishop, 2001), which, consequently, enhance academic achievement. Goodman and Hambleton (2005) concur with Bishop (2001) when they state test results provide a standardized way to identify students who may need additional support to gain the skills and knowledge all students should possess. Eventually, teachers will offer extra support to those students who are not able to pass the tests or assessments (Goodman & Hambleton, 2005). Walberg (2011) argues standardized testing contributes to the learning experience of students by offering them information on how well they have mastered the material intended for learning. Computers provide immediate feedback about correct and incorrect responses far more quickly and accurately than teachers and tutors can provide (Walberg, 20011). With this feedback, they can learn from their mistakes which boosts the learning process. Walberg (2011) also states students who are required to pass a standardized test for high school graduation learn more science and math. Phelps (2005) concurs students work harder and learn more as a result of the test. It encourages them to concentrate on meeting



standards and monitor their own progress (Walberg, 2011); such skills are important not only for increasing achievement but also for increasing success in life (Walberg, 2011).

Effectiveness of Standardized Tests as a Tool to Measure Educational Achievement

According to Kohn (2000) standardized tests provide a quick and easy method to measure progress but it can be inaccurate and biased. He also states we tend to believe numbers and, therefore, consider the test scores as fact. However, if we take an in depth look on how those tests are administered we will find the target was to find the easiest and fastest way rather than the most accurate. In other words it is much easier to assess how well we are performing something than to ask whether what we are doing makes sense (Kohn, 2000). Other scholars like Chomsky and Robichaud (2014) also believe there is bias detected through a thoughtful look on what is behind the scores, which lead to the conclusion they cannot be considered an accurate measures to evaluate academic performance. Chomsky and Robichaud (2014) believes standardized testing is based on the principle any person can be educated and tested under the same conditions. It is outlined by specific and external groups of specialists and politicians and claims to be fair and adjusted for all students, taking into consideration the ethnic and socioeconomic differences (Chomsky and Robichaud, 2014). To prove reality is different, they used the example of one particular question from a 6th grade science test asks students what is not a fruit. The choices are: A) orange, B) pumpkin, C) apple, D) celery. As any child who has encountered celery knows the answer is clearly choice D. But what if the parents of some children did not have the money to buy celery at the store and, therefore, those children do not know the appearance of celery (Chomsky & Robichaud, 2014). Goodman and Hambleton (2005) disagree with such claims and state there is simply little or no evidence to say exam bias is a serious problem. They suggest exam writers and reviewers put a lot of effort to identify and eliminate problems related to test bias (Goodman & Hambleton, 2005). Popham (2014) concurs standardized tests are carefully developed and institutions who administer them assign a large number of editors for that purpose.

There is also major criticism that many factors affect the test scores in addition to teachers' performance such as test anxiety. For some children test anxiety is a challenge, thus



their scores do not precisely reflect what they know, while others do not take the test seriously. For example, some students fill the multiple choice bubble sheet to resemble pictures of trees or cars. Socioeconomic situation of children also affect test scores (Kohn, 2000). Kohn (2000) states it is difficult to hold a teacher accountable for students' test scores when those scores represent all that has happened to the children before they arrived to the classroom. Such criticism makes it difficult to conclude if the teachers or the schools are the only variables affecting student achievement.

On the other hand, other scholars claim there is a misunderstanding about the purpose and the use of standardized testing when some researchers claim tests are too simple or too biased and stress such claims are naïve or deliberately misleading (Walberg, 2011). Walberg (2011) states research and experience prove standardized tests are generally good at measuring students' knowledge, skills, and understandings because they are objective, fair, efficient, and comprehensive. He concludes for these reasons they are utilized by decision makers with respect to admissions to colleges, graduate programs, and professional schools as well as qualification and licensing for many skilled occupations and demanding professions such as law and medicine (Walberg, 2011). They enable educators to evaluate the progress of students, identify their strengths and weaknesses, and plan remediation as well as updating of teaching and curriculum. Standardized tests grant parents the opportunity to monitor their children's learning and allow education leaders to monitor and evaluate the progress of schools and individual teachers.

Ethical Issues Related to Evaluating Teachers Based on Test Scores

Standardized testing is also causing fear, shame, anger and mistrust among some teachers (Segall, 2003). Ravitch (2010) stresses focusing on accountability in standardized tests produced fear among educators and the main goal of such tests has shifted from improving standards to implementing accountability. She concludes excessive focus on accountability threatens larger goals of education (Ravitch, 2010). Ravitch (2010) states the problem lies in the misuse of test scores for high-stakes decisions like rewarding or terminating teachers and in some cases closing schools down as standardized tests are not accurate instruments to measure achievement. The goal of testing as a formative tool to improve student achievement has changed to focusing on



evaluating teachers using test scores (Lee, 2011). However, Walberg (2011) opposed this approach and states effective schools focus on student learning, not on the satisfaction of the professional staff. If the data shows testing benefits students, it should be processed even if there is not consistent teacher support. Educators should take pride in seeing good results from their work and eventually get paid for the good performance of their students (Walberg, 2011). He stresses few educators and administrators oppose tests because they do not want the substandard results of their work to become better known and refuse to admit to themselves their teaching or management methods need improvement (Walberg, 2011). Many such educators are responsible for the cheating and fraud that has become prevalent in some school systems (Walberg, 2011).

Another issue is connecting teacher evaluation and sanctions to test score results can demotivate teachers to working with students who have special needs as the test scores will not reflect the time and effort invested to support children with special needs. Surveys have found teacher negative attitude towards neediest students have been associated with test based accountability particularly in high-need schools (Baker, et al., 2010). Ravitch (2010) thinks many principals limit the admission of struggling students because they negatively affect the school's test score in standardized tests. They may also encourage those students to stay at home or suspend them on the test day (Ravitch, 2010). Kohn (2000) believes competition over test scores is harmful and makes some teachers regard other teachers as obstacles of their own success with all the envy and hostility involved; therefore, individual teacher rewards based on comparing test results can also prevent teacher collaboration. For example, a teacher can hesitate to help students from another class on a mathematics problem if they thinks such help can improve class scores, thus his or her class will lose the competition. The ideal situation is to have all teachers collaborating for the sake of their students. Effective schools are collaborative institutions where teachers work across classroom and grade-level boundaries toward the common goal of educating all children to their maximum potential (Baker, et al., 2010). A school will be more effective if the teachers are more knowledgeable about all students and can coordinate efforts to meet students' needs (Baker, et al., 2010). Education is about learning instead of triumphing over everyone else (Kohn, 2000).



Implementing an evaluation system that depends on rewards and sanctions is likely to push talented teachers to avoid high-needs students and schools, or to leave the profession entirely (Baker, et al., 2010). A real life example of this approach is when the United States and Great Britain governments attempted to rank cardiac surgeons according to the survival rates of their patients. This resulted in surgeons turning away the most ailing patients (Baker, et al., 2010).

Comparing Educational Achievements of Different Countries

The results of international standardized tests are being used in the context of comparing the achievements of the educational policies in different countries. Governments of countries study the results and reflect on them in order to improve their education systems. Some countries as a result have been reforming their educational policies like France, Norway, and Switzerland (Baird, et al., 2011). In France, the government has been using the Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA) results to justify system reforms especially in primary school curriculum. Educational authorities in the country responded to PISA 2009 results by putting a plan of action including launching an initiative in science and introducing personalized learning assistance to help struggling students (Baird, et al., 2011). In Norway, the PISA results of 2000 and 2003 were below the Organization for Economic Cooperation and Development (OECD) average which shocked the country and triggered plans to improve curriculum and assessment (Baird, et al., 2011). In Switzerland, the results of TIMSS and PISA led to harmonizing curricula and standards across the country or at least within regions that speak the same language (Baird, et al., 2011). However, in other cases this process sometimes transforms into huge pressure on teachers to improve scores and results in merely teaching to the test. Chomsky and Robichaud (2014) stated the USA guiding principles of education are: competition and accountability. Finland has the singular goal of developing the humanity of each child and the Finnish educators do not understand the benefits of competition among teachers who should be collaborating for the sake of the students (Chomsky & Robichaud, 2014). Surprisingly, Finland has always achieved higher than the USA in international standardized testing like TIMSS though they do not believe in excessive testing and competition. For example, the final assessment of basic education in the



national curriculum requires student work samples of the last two years (Hendrickson, 2012) which can be considered as an interesting and more representative form of assessment to high-stakes testing. Some research studies concurred comparison of scores across nations can be misleading as countries have extremely different population size, first and second language, and cultural values (Di Giacomo, Fishbein, & Buckley, 2013).

Walberg (2011) thinks such comparisons are healthy as they spark competency and explained how American students spend about half the total study time Asian students do in regular schools and on homework. He concluded this is a major reason for their poor performance relative to Asian and European students in international surveys (Walberg, 2011). He believes the USA's poor achievement progress compared to other nations shows substantial improvements in teaching and learning are needed. Walberg (2011) believes competition drives innovation and productivity and could be successfully applied to the failing American school system. He gives an example on how competition in the technology market between Apple and Google's Android results in better services and similarly standardized tests empower families to choose the best school for their children just as consumers choose the best cell phone on the market (Walberg, 2011).

International Standardized Tests (PISA and TIMSS)

International assessments of educational achievement began in the 1960s and acquired attention in the 1990s by the public and governments because of broad media coverage (Beaton, Postlethwaite, Ross, Spearritt, & Wolf, 1999). Many countries are participating in two large-scale international assessments; TIMSS and PISA (Beaton, Postlethwaite, Ross, Spearritt, & Wolf, 1999). This research focuses on TIMSS. However, an overview of both assessments is provided as they are major international studies with circa seventy participating countries (Di Giacomo, Fishbein, & Buckley, 2013) and this could help to yield a deeper understanding of international standardized testing.

Overview of TIMSS

TIMSS is an international assessment which is used worldwide to evaluate educational reforms related to mathematics and science (IEA Mission Statement, 2015). TIMSS has been



running a four-year survey cycle since 1995 and is administered by the International Association for the Evaluation of Educational Achievement (IEA) which is an independent, international cooperative of national research institutions and governmental research agencies (IEA Mission Statement, 2015). It is an international assessment producing comparative studies of educational progress with the aim of deeply understanding the effects of policies and practices within and across systems of education (IEA Mission Statement, 2015). Through its comparative research and assessment projects, IEA's goals are to provide international benchmarks to assist governments in identifying the relative strengths and weaknesses of their education system, provide high-quality data to increase educational authorities understanding of key school and outside school-based factors which influence teaching and learning to prepare and evaluate educational reforms (IEA Mission Statement, 2015). IEA goals also include developing and improving the capacity of curriculum and enhance educational monitoring which would contribute to the development of a worldwide community of researchers in educational evaluation (IEA Official Website, 2015). TIMSS continues to be commonly used worldwide as an indicator of mathematics and science achievement and function of curricula (Di Giacomo, Fishbein, & Buckley, 2013). It assesses student knowledge in mathematics and science at Grades 4 and 8. As we understand from the mission statement, it is a curriculum related assessment. In addition, IEA claims to be an effective tool in evaluating achievements of different educational policies and eventually improving such policies by reflecting on the results. This supports the idea standardized tests is a necessary platform for measuring educational skills and developing policies to achieve excellence and competency among children.

Overview of PISA

PISA is an international assessment administered by The Organization for Economic Cooperation and Development (OECD). The mission of OECD is to encourage policies advancing the economic and social well-being of people around the world (About OECD, 2015). It is claimed to respond to the need for cross-national comparisons on student performance (About OECD, 2015). It is further posited it provides reliable information on how well education systems prepare students for further study, careers, and life. PISA also provides a basis for international collaboration in order to define and implement effective educational policies



(Organization for Economic Cooperation and Development, 2015). OECD (2015) argues PISA develops tests which are not directly related to the curriculum. Their tests are designed to evaluate how students at the end of compulsory education can use their knowledge in real life situations and be ready for full participation in society (Organization for Economic Cooperation and Development, 2015). Fifteen-year-old students from randomly selected schools worldwide take tests in the key subjects: reading, mathematics and science, with a focus on one subject in each year of assessment. In 2012, some countries also participated in the optional assessments of problem solving and financial literacy (Organization for Economic Cooperation and Development, 2015). The test lasts 2 hours and is a combination of open-ended and multiple-choice questions that are organized in groups based on a real-life situation. The students and their school principals also answer questionnaires to give information about the students' backgrounds, schools, and learning experiences and about the school system and learning environment (About PISA, 2015).

The UAE Perspective on Standardized Tests

The Knowledge and Human Development Authority (KHDA) in Dubai believes participating in international assessments has been an important step for the education in the Emirate with big benefits that can be seen throughout the school system (Knowledge and Human Development Authority, 2012). Motivated by the value of evaluating what students in Dubai could do compared to their peers regionally and internationally, policy makers have decided Dubai is now a participant in all major international assessments (Knowledge and Human Development Authority, 2012). The KHDA has an understanding the benchmark for success can no more be measured only by national standards. It is rather measured by knowing how students and education systems perform on an internationally comparable scale (Knowledge and Human Development Authority, 2012).

The United Arab Emirates (UAE) participated in TIMSS and PISA for the first time in 2007 and 2009 respectively. PISA is considered to have value in collecting important background information as all students complete background questionnaire when they take the test (Knowledge and Human Development Authority, 2013). Pairing such data with the proficiency data can be utilized by decision makers to establish standards for improvements. The



authority believes PISA represents a form of summative assessment which is common to the majority of students in Dubai. However, KHDA realizes PISA cannot be the ultimate tool to measure educational progress and curricula in Dubai. They stated while the OECD scale is created to measure trends, a simple change in a country or region's mean result does not automatically imply its 15-year-old students have developed additional skills (Knowledge and Human Development Authority, 2013). Over a three year period, countries can experience demographic changes. This is true in a dynamic region such as Dubai where a large expatriate community resides. To evaluate real improvement in students' readiness for life at age 15, further analysis, while taking into consideration such demographic changes, is required (Knowledge and Human Development Authority, 2013). Therefore, Progress can be measured when unrelated changes to the education system are considered (Knowledge and Human Development Authority, 2013).

There are also the student-level factors affecting achievement levels like attitude towards learning, confidence and self-belief. Nevertheless, the KHDA affirms international research constantly finds teachers to have the greatest effect on achievement through their instructional strategies to engage students in learning (Knowledge and Human Development Authority, 2013). Mr. Marwan Al Suwaleh, The UAE Ministry of Education Undersecretary, stated currently the country have no assessment of the number of high quality teachers in Emirati schools. In public schools, 83% of students in fourth grade were evaluated as below satisfactory levels in mathematics and reading, and similarly, 79% in science. In eighth grade, 75% of students scored below satisfactory levels in mathematics, and 71% in science. He added the ministry is not satisfied with the results (UAE Schools Should be up with The World's Best, 2014). So again there is an assumed relationship between test scores and teachers. TIMSS placed the UAE 23rd out of 42 countries (UAE Schools Should be up with The World's Best, 2014).

As we learn from the research literature many governments around the world give high importance to standardized tests to evaluate their policies and initiate reforms in countries like France, Norway, and Switzerland (Baird, et al., 2011). They also assess teachers' and curriculum's efficiency by such tests sometimes. This strategy has evolved from the need to have a measuring tool that adheres to international standards and reflects what students on a certain



age need to know. However, some research studies argue those tests follow the one size fits all approach which contradicts the principle of education. They affirm students have different learning styles and a lot of external factors can affect their achievements. Therefore, scores of such tests do not represent students' knowledge and skills. Some research studies advocate for other forms of assessment like work samples of students which are more representing than a single exam score.

Methodology

Research Design

This research study is exploratory, narrative, and interview based. The main goal was to explore and learn about standardized testing with a hope the findings and recommendations will be helpful for policy makers trying to make decisions depending on such tests. It is not intended to formulate an independent theory. The benchmark for inclusion in my literature review is how relevant and recent the material is to the pros and cons of standardized testing. The research scope is based on literature published between 1999 and 2015. Merriam (2009) defined the qualitative researcher as the one who focuses on learning how people explain their experiences and what implications result from such experiences. The interest is to learn how educators and educational institutions adjust to standardized tests and how they think about it rather than finding out how many students participated in a certain international test or conduct comparisons about achievements among countries who participated in it. Literature that supports and criticize standardized testing is included in order to keep high level of objectivity. It addresses three domains in my research: the debate on standardized testing, international standardized assessment and the United Arab Emirates (UAE) perspective on international standardized testing.

Data Collection Methods

This study uses books, journal articles, documents and interviews as instruments to collect data.



Interviews

Interviews are very useful and provide an insight into one's personal beliefs and thoughts. Merriam (2009) states interviewing is very important to observe behavior and feelings. Further, she considers interviews in qualitative research to be not rigidly structured and have more open ended questions. This study involves semi structured interviews with five teachers who had to prepare their students for the TIMSS assessment in Dubai in 2015. Their experiences have enriched the study. Moreover, TIMSS exam is a good example of a well acknowledged standardized testing.

Books, Journal Articles, and Documents

Literature was collected from ERIC and EBSCOhost databases and from the library of the American University in Dubai. In addition, google scholar was used to search for articles and books that discuss my research domains to obtain copies and collect the necessary data. Books and articles related to the three domains of my study were obtained either online or by purchasing printed copies. The search terms used were: pros and cons, benefits, teachers' perceptions, effects, problems and criticism. All of these terms were followed by standardized testing in the search box. Reference lists of relevant literature provided me with other potential sources. Data collected from such sources contributed to my understanding of pros and cons of standardized testing and the experience of some countries in that field. The KHDA report about Dubai TIMSS and PIRLS in 2011 was retrieved from the authority publications on the web and used to reflect the official opinion on the topic.

Sample Selection

Five grade four teachers were interviewed from two different schools. Pseudonyms were used to refer to the participants. Dana and Carla are mathematics teachers in School A. Nada, Ruba and Jacob are mathematics teachers in School B. Nada is also the primary mathematics coordinator in School B. School A follows its own prescribed international curriculum while School B defines itself as a school that follows the American system. The target was to get diverse perspectives from two different schools who participated in TIMSS last year so the results could contribute to the discussion and cover a spectrum of experiences and opinions.



Recording Data

Interviews were recorded using the voice recorder and transcribed. Notes were also taken through the interviews to record feelings, expressions, or behaviors that gave more meaning to what was being said. Interviewees were assured confidentiality would be maintained as the school name or their personal information would not be used in any way in my research. For data collected from readings, an excel sheet was designed to organize the themes that had emerged.

Data Analysis

Qualitative data collected from instruments was coded and analyzed afterwards. After conducting the interviews, another domain was added to the excel worksheet for the most significant statements said by the teachers. Statements supporting the standardized testing were placed beside the pros domain and the ones against it were placed beside the cons domain. This allowed a comparison of reading with teachers' experiences, therefore, validating the data collected. Recognizing analogy and differences helped to categorize and brief the information. Microsoft Excel helped to notice frequencies of some terms and statements which contributed to the general understanding of the issues related to standardized testing.

Reliability and Validity of Data

To increase the reliability and validity of the data, multiple resources were used to collect it which include interviews with teachers from two different schools who participated in TIMSS and the Dubai TIMSS and PIRLS 2011 Report. Interviewing teachers who work in two different curricula can enrich the data and enhance the results of analyzing it. Combining the data from the interviews with the findings of Dubai TIMSS and PIRLS 2011 report improves the validity of the data and yield a deeper understanding of the topic. An array of studies and research conducted by well acknowledged scholars who have an impressive career in educational research was also used in order to learn about the experience of other countries in standardized testing. To enhance the trustworthiness of the study all the resources are available upon request to the first author.



Results

This study set out to investigate the global use of standardized testing with a focus on the UAE experience in utilizing TIMSS. To this end, a thorough literature review was conducted to provide an understanding of how the world is currently focused on standardized testing and the resulting lessons learned. In addition, five interviews were conducted to obtain insight into the UAE perspective regarding the use of standardized tests, specifically the TIMSS. The findings from each interview will be presented accordingly. A document analysis of the standardized test administered in the UAE will follow.

Dana (School A)

Dana used to be the Mathematics coordinator for the primary section in School A in 2015-2016 academic year. She immigrated to Australia last June. However, she has returned back to Dubai recently as her husband had to move back to the city due to his job requirements.

Dana confirmed preparing for the TIMSS assessment was a difficult experience for students in her school, *“it was quite difficult for the students to adapt to the new way of thinking that depends on analysis”*. However, she said the addressed topics are important as they help students develop visual analysis and critical thinking skills especially in Geometry as this subject is not included in the curriculum of grade four in her school. Dana believes students did not really benefit from the experience, *“as we were teaching to the test and our main goal was to secure high scores”*. She was holding remedial classes for struggling students and thinks students did not show deep understanding of the taught concepts and thought of them as mere formulas with some applications. She believes this is due to the fact the new concepts were not given enough time in class as teachers had a short time to prepare for the assessment and that material was not included in the pacing chart of the school which was adjusted later to induce the new topics. Dana believes the result of such strategies is that students were memorizing answers instead of deeply understanding the concepts, *“I really believe that if enough time was given, then students would have learned more especially in geometry. Some students were merely memorizing answers which does not work in such international assessment”*. She stressed teachers were not given the results of the TIMSS exam but the administration included the



assessment's topics in the internal exams of the school and used their results to evaluate teachers and students.

As a coordinator Dana affirmed she used the results of those exams to assess teachers, *"as a head of department, I had to report some teachers to the administration as they did not perform well. On the other hand, I commended other teachers for their substantial efforts to adapt to the challenge of TIMSS"*. Dana explained how the students' averages were negatively affected in general due to the change of questioning methods, *"concerning students, the scores of many of them were negatively affected. Some of them really struggled while the others had minor issues. Only little number of them kept their usual performance or improved"*.

Dana further explained students in School A are used to direct applications and were not trained to make connections with other concepts in mathematics, real life problems or concepts from other disciplines. She stated students could perform well in TIMSS even without being prepared by their teachers if the curriculum matched the assessment's topics and children and teaching strategies infuse critical thinking as the way of learning.

When asked about the positives and negatives of standardized international assessments Dana replied that one of the positives is discovering gaps in curriculum and instruction. However, she affirmed the scores of such exams might not truly represent the knowledge of the learner as a student might be having a bad day on the date of the exam, *"a lot of factors can play a role in his score like family problems for example. Therefore, the result does not really reflect what he knows. I believe that the evaluation should include multiple assessments and not one single exam"*. Dana continued to explain one vivid example she can think of is her daughter who faced difficulties in performing well in writing after she joined her school in Australia. The problem was she was not well trained in writing in her previous school in Dubai. Eventually, her writing skills gradually improved in the new school and the teacher could see a significant progress. However, she did not do well in one of the assessments, *"she did not sleep well on the previous day which affected her concentration on the day of the exam"*. The child's teacher noticed the score did not really reflect her progress, *"she decided to evaluate her according to her work in the whole term instead of one single exam"*.



Dana is a strong believer of conducting multiple assessments to evaluate educational achievements, *“I believe that students’ classwork could be used as a form of assessment but without telling the students that we are evaluating them”*. Dana thinks the stress and fear associated with the word “exam” would affect the students’ performance and therefore the results would not be accurately reflecting their knowledge and skills. She said pupils, *“go to the exam with so many things to worry about like satisfying their teacher and parents with a good mark”*. Consequently, *“the pressure automatically increases and gets the child into a vicious cycle of stress and low scores”*. Dana concluded qualified teachers can evaluate their students according to their classwork.

Dana disagrees with the concept of comparing educational achievements of different countries as many of them have different curricula. She said it would not be fair to evaluate teachers and students according to scores when there are others factors that could play a role in the achievement like curriculum. She believes when standards are unified across the world then achievement could be linked to instruction and students’ work rather than curriculum variance. Dana said: *“I believe that standards should be unified before holding comparisons. For example, when I was a head of department I was evaluating teachers according to their students’ scores. She also believes all teachers in her school are, “teaching the same material so the only variable is the teacher and hence the results reflect his/her effort”*. Dana further explained how results could be affected by the different designs of curricula, *“in the system I used to work in, geometry is introduced in the secondary school while in other systems students start learning it in earlier stages”*. Dana said their pacing chart got changed to place significant importance on geometry before the TIMSS exam.

Dana concluded the interview by confirming to topics included in the TIMSS assessment are rich and rigorous. She wished such material is induced in the curriculum of the school from the beginning of the year rather than doing that just before the exam which deprives students from the opportunity to have a rich and joyful learning experience.



Carla (School A)

Carla taught mathematics in the primary section of School A for more than 3 years including the year of the last TIMSS assessment. She is currently teaching computing.

She believes preparing students for the TIMSS exam introduced students to critical thinking and real life problems which was not widely available in the curriculum of the school before. Carla was concerned some topics in TIMSS assessment did not match the curriculum and that is why the pacing chart of the school changed in order to induce the material required for the test. She said such a change was not smooth as, *“students got confused in the beginning and their scores significantly went down because they are not used to such type of questions that involve higher order thinking”*. Carla sadly stated, *“students’ averages were greatly affected. It was not easy for teachers at all. The shift from direct answer questions to questions that require the use critical thinking skills imposed a serious challenge”*. She believes such challenges are among the negatives of international standardized testing as, *“there was not enough time to gradually lead students through that shift and administration was very demanding regarding their scores”*.

However, she said she agrees with the questioning methods used in the international assessment as students should learn how to think rather than memorizing and automatically answering questions. She is confident any, *“educational system should train students on the skills of such questions that include critical thinking starting at grade 3”* instead of practicing them just before the exam which puts pressure on students and teachers. It appears Carla believes TIMSS exam measures critical thinking and the ability to apply knowledge in real life. She thinks the advantages of similar international assessments include pushing students to use their analytical skills and encouraging them to study more.

On the other hand, Carla is a strong believer in other effective methods of evaluation for educational achievement like project based assessment, *“I believe in project based learning. It encourages students to search for information and feeds their curiosity. They also feel that they are involved in their own learning”*.



Carla does not believe in comparing results of the countries after the assessment, “*as there are different cultures around the world and students come from different backgrounds so comparisons of achievements would not be accurate*”. She also thinks not all educational systems provide the same support for all their students so comparing students’ capabilities would be unfair. In her opinion, “*poor or developing countries cannot provide access to the internet for all students and hence rich countries would have advantage as their students’ knowledge will be richer due to their ability to research any information on the web*”.

Carla concluded the interview by suggesting to merge the material of the TIMSS assessment with the curriculum of the school so students and teachers can be equipped enough to succeed in such assessment.

Nada (School B)

Nada has been working in School B for five years as a mathematics and science coordinator for the primary section.

She stated the school’s participation in TIMSS was supposed to improve the learning experience of children but it turned out to be a shallow teaching to the test over a period of six weeks. She was agitated while telling me there was not enough time to cover newly introduced topics which were not originally in the pacing chart. Students’ averages were negatively affected and teachers were evaluated accordingly. Nada found the, “*material was difficult and teachers found it really challenging to teach*”. She exclaimed: “*rotational symmetry in grade 4! They are too young for that! Geometry was also difficult*”.

Contrary to the teachers interviewed previously, she believes students who depend on memorization would outperform other students in TIMSS. Nada finds the world’s focus on standardized testing to juxtapose differentiated learning, “*it’s ironic that the push for standardized assessments and the drive for differentiated instruction are growing at the same time*”. She believes assessments are necessary to evaluate achievement and policies but we should not depend on one single assessment as one size does not fit all. According to her we should acknowledge that, “*our students learn in different ways so they should also have different*



methods to demonstrate their knowledge”. She suggested, for example, giving an audio test for audio visual learners instead of a written assessment.

Nada believes in comparing educational achievements of different countries as this allows the exploration of emerging trends in education and inform policy makers of the gaps in order to adjust policies and curriculum accordingly. However, she admits, *“cultural or societal influences can impact students learning and achievement”*.

Nada stated the KHDA is pushing for good results in international assessments and this was clear when they told her school it will be partially evaluated according to TIMSS results. She said when averages of the international assessments are low, the KHDA is questioning why the scores are low while the averages of the school’s internal examination are high. This query took place after the last Iowa examination. Iowa is a standardized test developed by the University of Iowa in the USA and claims to inform educators about gaps in instruction and policy makers about the quality of educational achievements in order to support accountability (Iowa Testing Programs, 2016). Nada had to respond by saying they just integrated the material of TIMSS and Iowa exam in their curriculum and therefore it will take at least two years to see positive effects in the results of international assessment.

Nada was frustrated over the fact they have to teach their own curriculum in addition to the TIMSS and Iowa related topics which confuses the children and significantly increases the workload on teachers not mentioning how challenging it is to find the sufficient time for teaching all this material.

Ruba (School B)

Ruba has been teaching mathematics in School B for five years and lately holds remedial classes for struggling students after school. She has a different opinion compared to her colleagues in School B and the other teachers in School A. She believes the, *“students’ experience with TIMSS examination was rich and made them motivated and more focused on their learning”*. Ruba concurred with her colleagues in School B and teachers in School A some of the material required for the TIMSS exam is not included in the curriculum of the school.



However, she believes such a challenge helped the children to improve their skills and enhances the learning outcomes.

She explained the results were never shared with the teachers but she believes international assessments could help her to discover the gaps in the knowledge of her students. Ruba thinks such evaluation tool conforms to the increasing international focus on using standardized testing to evaluate the academic achievement of teachers, students, schools and countries.

Ruba believes, “international examinations add considerable pressure on teachers in terms of planning and preparation” as they need to attend to their students’ needs, school standards and exam requirements at the same time. In addition, finding time to teach the school’s curriculum as well as material related to the standardized tests, which is not included in the curriculum, is quite challenging. However, she believes such pressure is healthy as it exposes the students to themes and topics from different educational systems and improves their learning performance.

Ruba sounded as a strong believer in international assessment as she did not suggest any alternatives to measure educational achievement. She also affirmed, “*comparing the scores of different countries help keeping the standards high in spite of the fact that different countries have different educational policies and infrastructure*”.

Ruba suggested, “*training teachers on how to prepare their students for international examination in addition to inducing the required material in the curriculum of the school*”. She further explained some teachers know the material but do not have the skills to simplify it to their students or guide them on how to think and answer the questions in the exam.

Jacob (School B)

Jacob has been teaching and planning science lessons for other teachers in School B for two years. Prior to his appointment in School B, he taught in another school for three years.

Jacob does not believe participation in TIMSS exam last year reshaped the learning of his students. He thinks as a teacher he was merely following instructions and stressed, “*it was chaos.*”



We had to push the students to absorb a lot of information in a short period of time that consists of six weeks”.

He continued to explain their, *“curriculum is not aligned with the TIMSS material at all”*. For example, their science curriculum is guided by the Next Generation Science Standards (NGSS) which does not include many topics evaluated by TIMSS. NGSS is a framework of standards for science education that is developed in the USA and widely used by the states to prepare students for college and careers in addition to providing international benchmark for science literacy (Next Generation Science, 2016). He said the, *“NGSS does not cover the topic of force and motion magnets in grade 4 for example, while international examinations like TIMSS will ask questions about it”*. While exploring TIMSS related topics, the school downloaded some sample questions for TIMSS from the internet in order to have an idea about the tested units. Consequently, the administration had to change the lesson plans and pacing charts to address TIMSS material. Jacob was frustrated over the fact, *“teachers were merely drilling a lot of concepts for the exam and the learning experience was not student centered anymore”*. The main goal was to prepare students for the test rather than real learning experience.

He also confirmed results of TIMSS 2015 were not shared with the teachers though such results could give good picture about the students’ educational attainment. Jacob believes TIMSS is supposed to measure critical thinking, analytical skills and knowledge students gained throughout their school years until the year of the examination. However, he thinks some students have problems with retaining such a large amount of information and skills so such comprehensive testing would be a problematic for them.

Jacob said the pressure associated with such tests was tremendous. *“The school doubled the Mathematics and Science periods for the students. In order to do that, the administration had to reduce English, Social Studies, Physical Education and Art periods so they can make space for the additional Mathematics and Science periods in the schedule”*, thus placing extra pressure on teachers and students. Eventually the stress increased with the school’s daily and weekly assessments as students had to demonstrate their knowledge of a huge amount of material that was taught in a very short period of time. He insisted not all students have advanced retention



skills to be able to cope with the requirements. The scores of the internal examination tested the TIMSS topics were counted in the students' averages which exponentially heightened the tension for teachers and students.

Jacob agrees some students know more information after the TIMSS experience. He could see students motivated and more focused on their learning as they realized they should achieve positive results in international examinations. However, he believes more time should be given to learning the explained concepts so students can deeply understand them.

Jacob stated International assessments will not be needed to discover gaps if internal examination in any school is developed by professional teachers following high quality guides or references to edit exams.

He strongly disagrees with the concept of comparing educational achievement of different countries due to distinct, "*cultural backgrounds, mentalities, lifestyles and parental support that those countries have*". He said such comparisons could be held among different schools in the same country instead.

Jacob realizes external examinations are there to stay because of the world focus on standardized testing. He explained the KHDA always compares the results of internal examination in any school with its performance in international assessments like TIMSS and Iowa. They question the school about the reason for any variance between the results of the school and international assessments.

He also strongly believes a delicate and skillful integration of the TIMSS topics in the curriculum from the beginning of the year will smoothly prepare students for such testing instead of drilling the material just couple of weeks before the exam. According to him this would mean eliminating some topics in the NGSS that do not get evaluated in the TIMSS because, "*piling up topics from different systems in the curriculum would not work as there is not time to cover all the material from a mixture of educational systems*". Jacob affirmed once the decision is made to do such integration then the students would perform better in TIMSS. He believes students' performance in international examination is not excellent because, "*we do not have the culture of reading*". He stated the UAE's nationals and expatriates should be encouraged to read science



and mathematics books in addition to searching for information on the internet. In other words he believes educators should promote independent learning that broadens the acquired information in school.

Dubai TIMSS and PIRLS 2011 Report

In order to get an insight into the official perspective about standardized testing in the country, a copy of the Dubai TIMSS and PIRLS 2011 Report was obtained from the KHDA publications.

Dubai and UAE achievement against other countries

The report revealed there was a variance in achievement among students from schools following different curricula in Dubai (Knowledge and Human Development Authority, 2012). Other countries participated with a number of their schools while Dubai participated with all public and private schools that teach grade 4 or grade 8 and use Arabic or English as the language of instruction (Knowledge and Human Development Authority, 2012). A sample of 6443 students taught different curricula were tested in Dubai in 2011 (Knowledge and Human Development Authority, 2012). For grade 4 mathematics, Dubai scored 468 and the United Arab Emirates in general scored 434 which is below 500- the TIMSS scale center point (Knowledge and Human Development Authority, 2012). However, the emirate's score significantly exceeds the score of other MENA countries like Bahrain (436), Qatar (413), Saudi Arabia (410), Oman (385), Tunisia (359), Kuwait (342) and Morocco (335) (Knowledge and Human Development Authority, 2012). In grade 4 science, Dubai and the UAE in general scored 461 and 428 respectively and both scores are below the center point of TIMSS (Knowledge and Human Development Authority, 2012).

Strength and Weakness

Dubai's TIMSS and PIRLS 2011 report stated geometric shapes and measures occupied 35% of the content domains evaluated in the assessment. It also revealed the performance of grade 4 students in that section was significantly weaker than in other sections (Knowledge and Human Development Authority, 2012). Dubai grade 4 students showed small but compelling strength in the knowing content domain relative to mathematics overall (Knowledge and Human



Development Authority, 2012). However, they had statistically significant weakness in the reasoning domain (Knowledge and Human Development Authority, 2012).

Achievement according to Curricula

The report declared the UK and IB schools of Dubai achieved the highest scores on average (Knowledge and Human Development Authority, 2012). Students learning in US curriculum schools scored less than Indian curriculum schools but better than those in the Ministry of Education (MoE), Pakistani and Philippine curricula schools (Knowledge and Human Development Authority, 2012). Students of UK and IB curriculum schools showed their superiority at the advanced international benchmark level as well (Knowledge and Human Development Authority, 2012). This is a level at which students can perform complex analyses and reasoning. At least 11% of students at UK curriculum schools and 9% of students at IB curriculum schools achieved at that level (Knowledge and Human Development Authority, 2012). 58% of students in MoE public and private curriculum and 70% in the Pakistani curriculum demonstrated achievement levels below the low international benchmark which is an alarming level at which students can only show basic skills (Knowledge and Human Development Authority, 2012).

Effects of Results on Schools' Ratings

KHDA states international assessment is one of many sources of information about the quality of education students are receiving in Dubai and stresses the results of such assessments provide a useful guide to annual inspections (Knowledge and Human Development Authority, 2012). Such results inform inspection about students' achievement in a school against international standards (Knowledge and Human Development Authority, 2012). The report affirms TIMSS results appear to be closely related to inspectors' rating of school quality though they are not the only indicator (Knowledge and Human Development Authority, 2012). The authority justifies the use of such gauge by saying schools with outstanding and good inspection rating demonstrated high performance in TIMSS mathematics at the same time (Knowledge and Human Development Authority, 2012).



Evaluation and Reforms

The report concludes data collected by holding comparisons of achievements in international assessments such as TIMSS is an important tool to support decision-making and initiate evidence-based reforms (Knowledge and Human Development Authority, 2012).

Discussion and Conclusion

This Study explored the UAE perspective on standardized international testing through interviews and document analysis of the Dubai TIMSS and PIRLS 2011 report. Data was collected and analyzed qualitatively. Two major themes have emerged while interpreting the results: curriculum and learning Experience and academic evaluation which will provide the structure of this chapter.

Curriculum and Learning Experience

Among all interviews, the problem of curriculum kept on presenting itself. Obviously, interviewed teachers from both schools do not have their curricula aligned with the international assessments' standards and had to abide by their own curriculum in addition to teaching the TIMSS topics. Most of the teachers ended up having excessive material to teach in a short period of time and concurred this had disrupted the learning experience of their students.

In order to address the international assessment, the schools' pacing charts and lesson plans were changed just six to eight weeks before the date of the examination to focus on TIMSS material; therefore, teaching to the test was prevalent. In School B, the administration was desperately trying to find time for teaching the TIMSS material and one of the solutions was to substitute arts, music, social studies, physical education, and English periods for mathematics and science lessons. Not all of them would be mathematicians or scientists in the future as their interests vary and I believe their substitution violates the learning experience of students in those subjects. The removal of liberal arts from the curriculum could make students resent mathematics and science which potentially contradicts the goals of such assessments. The difficulty is naturally compounded when even, recess periods taken substituted by more work. Besides, such decisions do not yield an authentic learning experience for students as they are not sponges of information dispensed by teachers. Short time could compel teachers to explain



concepts quickly rather than allowing learners to discover and deeply understand those concepts. Students are supposed to construct knowledge and be able to synthesize learned concepts to develop transferable concepts instead of learning certain facts needed for a single test. In School A, critical thinking is not widely facilitated in the curriculum and getting students to shift to such a sophisticated way of learning in six weeks does not provide them with enough time to develop their higher order thinking skills needed for such assessment.

Studying the KHDA report about 2011 TIMSS and PIRLS assessments in Dubai reveals curriculum plays an important role in achieving better results. The report classified IB and British curricula schools to be achieving the best results in TIMSS. A further research on each curriculum can give a better insight into the reasons for their achievement. However, connecting the report's classification with the teachers' statements about their struggle due to the mismatch between their curricula and TIMSS can lead us to believe curriculum is a factor of success in the assessment. On the other hand, the KHDA could use the assessment to discover trends in the curricula across schools in Dubai in order to improve policy. The authority would look into components of IB and British schools' success in TIMSS and recommend inducing those components in other curricula. As discussed earlier in the literature review, TIMSS results provide an opportunity for governments to update curricula and improve policies (Bishop, 2001; Phelps, 2005; Walberg, 2011).

Teachers' responses indicated there is a major weakness in geometry. It seems geometry is not widely taught in grade four across Dubai schools. This weakness is confirmed by KHDA TIMSS and PIRLS, 2011, Dubai Report as it revealed geometry to be the greatest challenge for almost all grade four students across Dubai (Knowledge and Human Development Authority, 2012). Nevertheless, it is a major tested domain in TIMSS and constitutes 35% of the mathematics exam (Knowledge and Human Development Authority, 2012). Schools whose curriculum includes few or no geometry topics in grade 4 would be encouraged to reconsider their students' learning in order to improve student achievement on TIMSS. From my experience as a mathematics teacher, certain units in geometry like types and measurements of angles and lines, identifying certain quadrilaterals and triangles can be introduced in grade four. There are



also many other topics of the same level of difficulty and meet the cognitive level of students in grade four.

The ideal situation would be an inclusive curriculum for each school of the assessment's topics and, therefore, there is no need to teach to the test. Sooner or later, decisions about what is most important and what needs to be eradicated would have to be taken by the school's administration.

Academic Evaluation

We can learn from teachers' responses their experience with TIMSS exam was difficult. Their administration put pressure on them in order to raise the scores. Yet, those teachers were not provided with sufficient time to make the process meaningful. Most of the interviewees agreed changing the pacing charts just a couple of weeks before the exam to include the material related to the standardized assessments transformed teaching and learning into a transactional process. This concurs with Kohn's (2000) statement high scores sometimes represent excessive drilling to excel in a particular test rather than learning. I believe assessment should be a procedure of discovering gaps in the knowledge of students rather than assessing what teachers drilled in the last four to six weeks. It is true TIMSS and other international assessments help governments discover gaps in curriculum and policy at a national level and the results inform their decisions on improving education in their country (Bishop, 2001; Phelps, 2005; Walberg, 2011). However, there is a problem in the application of this concept as the pressure is being put on teachers in order to raise the scores and little attention is given to effects on the classroom. The KHDA's evaluation of schools is partially influenced by the schools' performance in TIMSS (Knowledge and Human Development Authority, 2012). Both schools are heavily dependent on teachers to meet that demand by evaluating their performance according to their results in the internal summative assessments inclusive of TIMSS material. Yet, four out of the five interviewed teachers insisted there was not enough time to properly teach the material so they ended up teaching to the test. According to some of the reviewed literature, teachers try to make their students absorb as much information as possible in a short time which kills the creativity in class (Chomsky & Robichaud, 2014; Darling-Hammond & Turnipseed, 2015). It shrinks the time



teachers can dedicate to meeting the various needs of their students, whereas the tests themselves favor students with certain learning styles while disadvantaging other styles (Olivant, 2015)

It is well researched students learn in different ways (Fleming & Baume, 2006) and have multiple intelligences (Gardner, 2006), so such practices limit the time for differentiation as teachers find themselves rushing to finish the required material. Ravitch (2010) highlighted the problem that lies in the misuse of test scores for high-stakes decisions like rewarding or firing teachers and in some cases closing schools down because standardized tests are not accurate instruments to measure achievement. The goal of testing as a formative tool to improve student achievement has changed to focusing on evaluating teachers using test scores (Kuehn 2010; Lee, 2011).

Teachers play a very important role in education and they should be provided with a certain support system in order to achieve better results. Such a system could include continuous professional development supported by a coherent, consistent and comprehensive curriculum that include topics of major assessments. Teachers need time to foster creative, critical thinking, and deep understanding of the core concepts in the classrooms rather than trying to fill the gaps between the curriculum and the international assessment weeks before the exam date. In other words, this is a process that takes place throughout the year instead of half a term. Schools can properly evaluate teachers once a support system is established which provides them with the necessary factors of success and eventually the results would reflect their performance.

Four out of the five interviewed teachers did not agree with the principle of comparing educational achievement of countries according to results of international standardized testing. They believe factors like cultural and financial background play a role in such testing. It is true each country has different priorities in education and standardized testing cannot precisely cater for all those diverse priorities. Di Giacomo, Fishbein, & Buckley (2013) concur comparison of scores across nations can be misleading, as countries have extremely different population size, first and second language and cultural values. However, there is a certain platform of standards in science and mathematics which are unarguably important and essential and should be global. Therefore, TIMSS can help authorities to discover if their policies and curriculum have met



those standards. For example, Walberg (2011) believes the USA's poor achievement progress compared to other nations shows substantial improvements in teaching and learning are needed.

Some teachers suggested other ways of assessment to replace, or add to standardized testing like projects and classwork. This concurs with Finland final assessment of basic education in the national curriculum which requires student work samples of the last two years (Hendrickson, 2012). Using such methods to enhance their knowledge and skills needed for standardized testing is a more reasonable suggestion. In other words, let us allow students to express their cognitive skills and what they know through their preferred methods. This will involve them in their own learning and intrinsically motivate them to construct deep and transferable knowledge about the intended content. Eventually, we can teach them how to express their understanding in any standardized testing which could be done over few sessions that prepare them for such task. However, this plan cannot be done in six weeks before the exam. As discussed previously in the curriculum section, once the curriculum is aligned with TIMSS topics then sufficient time will be available for students to learn and demonstrate knowledge of those topics in internal and international assessments. There are worries this will not meet local school needs.

It can be concluded it would be desirable for schools to consider aligning their curricula with the TIMSS topics. Many schools in Dubai follow different curricula, however, one of the targets of the UAE vision for 2021 is to rank among the top 15 countries in TIMSS as the country ranked 23 of 42 in 2011 cycle (First-Rate Education System, 2016). Schools are encouraged to embrace this vision and thrive to achieve this ambitious goal. One of the components of achievement is to integrate TIMSS material in the grade four curriculum and give it sufficient time to be taught and learnt. This integration includes geometry as an important subject that greatly contributes to raising the scores of the UAE in the assessment. Once the integration has taken place, then academic evaluation of the schools, teachers, and students' performance would be more fair and effective. All schools then will have the same benchmark standards and their results will depend on the quality of their work on those core standards.



This will have to go hand-in-hand with supporting teachers and provide them with appropriate professional development and well outlined curriculum, instead of solely holding them responsible for poor results. Accountability is important, but only when there is equity in providing success factors so individual teachers can represent their skills in utilizing those factors.

Finally, further study of the issue of standardized testing is desirable in order to have a deeper insight on how to improve the UAE's performance on international assessments. This study is limited to five teachers from two different schools, therefore, generalizability is limited due to sample size. However, it generated ideas about some of the difficulties and challenges related to international assessments and has the potential to inform larger scale studies using this research as a starting point.

Limitations

This is not a large scale study conducting extensive research in many schools in Dubai. The findings might not be applicable to all schools as the size and the population widely varies amongst them. In addition, some teachers or administrators might not feel comfortable talking about their experience with preparing their students for standardized tests.

Implications

The UAE began participation in the international surveys in 2007 which could be considered as a recent experience. We have a tendency to believe anything labelled as international is valid and has good quality. This study might remind educators objectivity should be maintained because there are positive and negative consequences in the macro and micro for the education system. Therefore, it is a hopeful attempt to assist educators in having an insight into the experience of other countries and the UAE with standardized tests and the possible issues that emerged after using them to evaluate educational achievement. This research might contribute to the discussion of the topic and could generate further and large scale studies in the UAE that can enhance our knowledge about standardized testing. It is possible for such studies to provide rich background for policy makers.



References

- Aydeniz, M., & Southerland, S. (2012). A National Survey of Middle and High School Science Teachers' Responses to Standardized Testing: Is Science Being Devalued in Schools? *Journal of Science Teacher Education*, 23(3), 233-257. Doi: 10.1007/s10972-012-9266-3.
- About OECD*. (2015). Retrieved from the Organization for Economic Co-operation and Development: <http://www.oecd.org/about/>.
- About PISA*. (2015). Retrieved from the Organization for Economic Co-operation and Development: <http://www.oecd.org/pisa/aboutpisa/>.
- Baird, J., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T., & Daugherty, R. (2011). Policy effects of PISA. *Report commissioned by Pearson UK*
<http://oucea.education.ox.ac.uk/wordpress/wpcontent/uploads/2011/10/Policy-Effects-of-PISA-OUCEA.pdf>.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. EPI Briefing Paper# 278. *Economic Policy Institute*.
- Beaton, A. E., Postlethwaite, T. N., Ross, K. N., Spearritt, D., & Wolf, R. M. (1999). *The Benefits and Limitations of International Educational Achievement Studies. Trends in Education*. International Institute for Educational Planning/UNESCO, 7-9 rue Eugene-Delacroix, 75116 Paris, France.
- Bishop, J. H. (2001). Using High Stakes Tests to Raise Achievement.
- Chomsky, N., & Robichaud, A. (2014). Standardized Testing as an Assault on Humanism and Critical Thinking in Education. *Radical Pedagogy*, 11(1), 3.
- Development Overview*. (2016). Retrieved from Next Generation Science Standards Website: <http://www.nextgenscience.org/development-overview>.



- Di Giacomo, F. T., Fishbein, B. G., & Buckley V. W. (2013). TRENDS.
- Dubai PISA 2012 Report. Knowledge and Human Development Authority. (2013). Retrieved from KHDA Publications: <https://www.khda.gov.ae/en/publications/article?id=160>
- Dubai TIMSS and PIRLS 2011 Report. Knowledge and Human Development Authority. (2012). Retrieved from KHDA Publications: <https://www.khda.gov.ae/en/publications/article?id=137>.
- First-Rate Education System*. (2016). Retrieved from UAE Vision 2021: <http://www.vision2021.ae/en/national-priority-areas/first-rate-education-system>
- Fleming, N., & Baume, D. (2006). *Learning Styles Again: VARKing up the right tree!*. *Educational Developments*, 7(4), 4
- Gardner, H. (2006). *Multiple intelligences: New horizons*. Basic books.
- Goodman, D., & Hambleton, R. K. (2005). Some Misconceptions About Large-Scale Educational Assessments. In R. P. Phelps, *Defending Standardized Testing* (pp. 91-108). Mahwah, New Jersey, USA: Lawrence Erlbaum Associates.
- Hendrickson, K. A. (2012). Assessment in Finland: A Scholarly Reflection on One Country's Use of Formative, Summative, and Evaluative Practices. *Mid-Western Educational Researcher*, 25(1/2), 33-43.
- IEA Mission Statement*. (2015). Retrieved from IEA Official Website: <http://www.iea.nl/?id=72>
- Iowa Testing Programs*. (2016). Retrieved from The University of Iowa: <http://itp.education.uiowa.edu/Default.aspx>
- Kohn, A. (2000). *The Case against Standardized Testing*. (L. Bridges, Ed.) Portsmouth, New Hampshire, USA: Heinemann.
- Kuehn, L. (2010). Standardized Testing is a Global Issue. *Our Schools / Our Selves*, 20(1), 69-75.



- Lee, L. (2011). What Did the Teachers Think? Teachers' Responses to the Use of Value-Added Modeling as a Tool for Evaluating Teacher Effectiveness. *Journal Of Urban Learning, Teaching, and Research*, 797-103
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.
- Olivant, K. F. (2015). "I Am Not a Format": Teachers' Experiences with Fostering Creativity in the Era of Accountability. *Journal of Research In Childhood Education*, 29(1), 115-129. doi:10.1080/02568543.2014.978920
- Organization for Economic Cooperation and Development. (2015). *Beyond PISA 2015: A Longer-Term Strategy of PISA*. Retrieved from Organization for Economic Cooperation and Development Official Website: <http://www.oecd.org/pisa/pisaproducts/Longer-term-strategy-of-PISA.pdf>
- Phelps, R. P. (2005). *Defending Standardized Testing*. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates.
- Popham, W. J. (2005). Standardized testing fails the exam.
- Popham, W. J. (2014). Making Sense Out of Standardized Test Scores. In W. J. Popham, *Classroom Assessment* (pp. 327-344). United States of America: Pearson
- Race to the Top*. (2015). Retrieved from The White House: <https://www.whitehouse.gov/issues/education/k-12/race-to-the-top>
- Ravitch, D. (2010). *The Death and Life of Great American System: How Testing and Choice are Undermining Education*. New York, United States of America: Basic Books
- Rizga, K., & Hernandez, K. (2015). Sorry, I'm Not Taking This Test. *Mother Jones*, 40(5), 38-62.
- Segall, A. (2003). Teachers' Perceptions of the Impact of State-Mandated Standardized Testing: The Michigan Educational Assessment Program (MEAP) as a Case Study of Consequences. *Theory and Research in Social Education*, 31(3), 287-325.



The International Association for the Evaluation of Educational achievement (2015, January).

IEA Policy Brief: Is Teacher Experience Associated with Mathematics Achievement?

Retrieved from IEA Official Website:

http://www.iea.nl/fileadmin/user_upload/Policy_Briefs/IEA_policy_brief_Aug2015.pdf.

Turnipseed, S., & Darling-Hammond, L. (2015). Accountability Is More Than a Test

Score. *Education Policy Analysis Archives*, 23(7-11), 1-8. doi:10.14507/epaa.v23.1986.

UAE Schools Should be up with the World's Best. (2014, September 18). Retrieved from The

National: <http://www.thenational.ae/uae/education/uae-schools-should-be-up-with-the-worlds-best>.

Walberg, H. J. (2011, May 20). *Stop the War Against Standardized Tests*. Retrieved from Hoover

Institution Official Website: <http://www.hoover.org/research/stop-war-against-standardized-tests>.

Walberg, H. J. (2011). *Tests, testing, and genuine school reform*. Hoover Press