

International Journal of Education Policy & Leadership

Saudi National Assessment of Educational Progress (SNAEP)

ABDULLAH SALEH AL SADAAWI
King Saud University

ABSTRACT: To provide a universal basic education, Saudi Arabia initially employed a rapid quantitative educational strategy, later developing a qualitative focus to improve standards of education delivery and quality of student outcomes. Despite generous resources provided for education, however, there is no national assessment system to provide statistical evidence on students' learning outcomes. Educators are querying the curricula and quality of delivery for Saudi education, especially following low student performances on the Trends in International Mathematics and Science Study (TIMSS) in 2003 and 2007. There is a growing demand for national assessment standards for all key subject areas to monitor students' learning progress. This study acknowledges extant research on this important topic and offers a strategy of national assessment to guide educational reform.

Al Sadaawi, A. S. (2010). Saudi National Assessment of Educational Progress (SNAEP). *International Journal of Education Policy and Leadership* 5(11). Retrieved [date] from www.ijep.org.

Background

As the largest country in the Middle East, Saudi Arabia occupies the greater proportion of the Arabian Peninsula, sharing its northern borders with Kuwait, Iraq, and Jordan, its southern border with Yemen, and its eastern reaches with Oman, the United Arab Emirates (U.A.E.), and Qatar (Siddiqui, 1996). The country is divided into 13 regions, or administrative divisions, and contains over 6,000 cities, towns, and villages (Al-Sadan, 2000). The population, according to the Central Department of Census and Information, was 22,673,538 in mid-2004, including some 6 million expatriates (Ministry of Economy and Planning, 2005).

The first Saudi educational system was established in 1924, initially as a few primary schools for boys (Al-Sadan, 2000). In 1938, the Directorate of Education assumed control of private schools, excluding military academies (Rugh, 2002). With oil rents flowing after the second world war, the country's founder King Ab-

dulaziz Al-Saud initiated an extensive educational framework incorporating six years of primary school followed by three years each of intermediate and secondary school (Ministry of Education, 2001). By 1951, the country had 226 schools with about 30,000 students. In 1953, the Ministry of Education replaced the Directorate of Education, specifically to improve the school system for male students to meet international standards, although the core of the system was, and still is, religious studies (Al-Sadan, 2000). The General Presidency for Girls' Education was established in 1960, responsible for developing a public education system to enable women to access knowledge and skills to become active members of society. The organization supervised all public girls' schools and education (Al Hakami, 2000). In 2003, the Ministry of Education (2005) assumed responsibility for girls' schools and integrated girls' schools into the public education system. As a consequence, Saudi Arabia has a higher rate of literacy, 85 percent, compared to other Arabic countries, such as Egypt and Algeria, both of which

developed their education systems prior to Saudi Arabia (United Nations Development Programme (UNDP), 2009).

Education Resources

Saudi Arabia has a substantial school system, with over 30,000 schools educating 5 million students (Ministry of Education, 2006). However, there is a vigorous debate regarding the substance of that education. The advocates for the traditional approach, which gives precedence to Arabic values, argue that the Saudi government promotes skills and knowledge acquisition as education for all, providing the citizens with free education from state primary school to public university graduation. Students have monthly remunerations and all study resources (Bahgat, 1999; Rugh, 2002).

Over 25 percent, Saudi Riyals (SR) 137 billion (U.S. \$35.5 billion.), of the country's 2010 budget was allocated to Education and Manpower Development "to increase both the quality and quantity of the Kingdom's human resource base among the Saudi Arabia's population, of which two-thirds are under the age of 30" (U.S.-Saudi Business Council, 2010). Further, overarching educational projects are funded off-budget; for example, the King Abdullah Public Education Development Project received SR 10 billion (U.S. \$2.7 billion).

Education Standards

Nevertheless, critics assert resources are focused on quantitative matters, such as numbers of facilities, teachers, and students, to the detriment of the curricula standards and quality of pedagogy (Abdaljoad, 2005; Al-Ebarahim, 2001; Al-Ghamdi, 2005). Schools' administrations may be compounding elements within the system that are adversely affecting students' learning outcomes. The thrust of resource allocation to date centers on the educational needs of young children, responding to the 3.2 fertility rate (UNDP, 2005). Nevertheless, outcomes from this quantitative approach are mixed, with inadequate schools and facilities to meet the numbers of school-age children. For example, as a temporary measure, the Ministry is renting private residences to use as schools. However, without a national strategy to allow resource flexibility, this makeshift measure has resulted in a profusion of

rented premises used for all education purposes, until they now exceed the number of government-owned and appropriately resourced school buildings. In 2006, there were 320 rented schools and 246 government schools in Riyadh alone (Ministry of Education, 2006). Intended for housing, the rented premises are inappropriate for schooling purposes, lacking dedicated and well-equipped classrooms, safety requirements, laboratories, and libraries (Al-Otaibi, 2005; Almegidi, 2004). The Ministry of Education attributes this situation to Saudi's population growth rate, which is relatively high, but does not exceed other Gulf countries that can plan their school resources effectively, such as the U.A.E. (UNDP, 2005).

The Ministry of Education, while challenged for accommodation, is improving standards of educational delivery and the quality of outcomes for students. Strategies are being formed to identify optimum paths toward skilled human resources to continue the country's economic and social development. At the time of this study, several key programs are being used to pilot different approaches in different educational areas. Examples of these projects follow, as an indication of the directions the Ministry is pursuing.

Saudi Leading Schools Project

This project is derived from the education system in Victoria, Australia. It was instituted in Saudi Arabia in 2002 as a means to address deficiencies in public schools' responses to individual students' needs. This model is interactive, technical, and incorporates a variety of resources. It accords the school a degree of self-administration, and provides a flexible curriculum based on openness and partnership with the local community (Ministry of Education, 2005). After six years of implementation in five primary and intermediate schools, the project is still under review.

Schools Evaluation Program

This program is a continuing and comprehensive evaluation process for all aspects of a school, including environment, administration, teachers, and students. It relies upon specific measurement tools and uses quantitative targets (Ministry of Education, 2004). The project is based on educational measurement and quality standards used by the British Office for Standards in Education, the entity that United Kingdom (U.K.)

makes responsible for evaluating public schools (Al-Hamed, 2005).

Schools Personnel Evaluation Project

A project to identify professional standards for teachers, supervisors, and school principals by using objective measurements began with a pilot for a teacher competency assessment in 2001 (Ministry of Education, 2004). The assessment is now one of a suite used for teacher selection, while supervisor and principal competency criteria were being evaluated at that time.

Thinking Skills Development Program

In 2002, a program was implemented to assist students to move from rote learning habits to developing problem solving skills and seeking creative solutions. This entails restructuring the primary school curriculum to introduce interactive class activities, and therefore implement improvements to cognitive skills for educational practitioners. The first phase of the project included a training program for educational leaders and issuing a teacher's guide for the development of thinking skills (Ministry of Education, 2004).

However, educators and researchers found that the Ministry's stand-alone projects were not focused on the country's education goals, they lacked integration or communication, and few have reported on their progress or application elsewhere. Further, most of the projects have been curtailed or abandoned during the last few years. For example, the General Project for Curriculum Development commenced in 1999 with a target completion of 2002, yet remained suspended in 2007, and the Saudi Leading Schools Project described above was marginalized after years of teacher training and experiment. Researchers attributed the absence of clarity in education development to two factors; first, there is no consensus of standards in the curricula; and second, an independent office is required to evaluate development projects and to oversee the Ministry of Education's performance (Abdajoad, 2005; Al-Baadi, 1995; Al-Daoud, 2004; Al-Dossary, 2000; Al Hakami, 2004).

Educational Outcomes

There is an emerging consensus among educators, researchers, and policymakers that the Saudi education system is not achieving best practices in its provision

of facilities, curricula, or teaching standards. Some time ago, the Ministry of Education Report 2000–2004 (2004) stated that the educational challenge facing the country was to prepare students for a competitive workforce; and to achieve this goal, emphasis on teaching quality mathematics, science, and technology was required through the adoption of best practice in teaching and learning methodologies.

Traditional teaching practices for some developing countries were based on rote learning in small groups, and the Saudi educational system initially followed this path, adding on curricula in a traditionalist setting. In the developed countries, curricula and teaching standards were adopted earlier, but the traditionalist approach in most developing countries persists. Thus, for the education authorities, there is the grave issue of their inability to develop a comprehensive and reliable school system from the considerable resources devoted to education over the last decades, and this issue is reflected in the uniform substandard results.

Administration of a country's judgmental goal, such as preparing future citizens for a future workforce, is not unique to Saudi Arabia, and issues cascading from the aspiration continue to fuel considerable research effort. However, there are fundamental steps to implementation of any program: motivation, goals, resources and, most importantly, standards. The absence of valid and reliable indicators to monitor Saudi educational outcomes over the decades is a substantial disincentive to achievement. Schools' development drifts; a successful school and an unsuccessful school share the same level of resources and general curricula, but neither report against targets. There is little responsibility shown by the entire educational system for the Ministry's goal: preparation of students for a competitive workforce. The TIMSS 2003 assessment, an international test of mathematics and science achievement of students from participating countries, provided an unwelcome result in the ranking of Saudi students. Saudi 8th grade students rank some 34 percent lower (331/500) than the international average in TIMSS mathematics test results, and 20 percent lower (399/500) in science results (Martin, Mullis, Gonzalez, & Chrostowski, 2004; Mullis, Martin, Gonzalez, & Chrostowski, 2004). For the 2007 tests, Saudi students' mathematics performance did not improve, at 33 percent below average (329/500); however, the science result improved to 19 percent below average

(403/500) (Martin, Mullis, & Foy, 2009; Mullis, Martin, & Foy, 2009).

Given that education commands about a quarter of the Saudi general budget, this is a serious and unexpected outcome from the international comparison of 8th graders' performances. Informed opinion (Fensham, 1998; Jakwerth, 1999; Kind, 1999) determined that a comprehensive and agreed-upon national assessment system to benchmark results, generate statistical data, and permit adequate comparison is essential to raise standards in curricula and delivery. This study aims to add to debate supporting national assessment standards to drive change and significantly improve performances in the school system. The next section discusses the recognized antecedents of national assessments.

National Assessments

National assessments provide statistical data on benchmarked school student performances to inform policymakers, assist curricula development and delivery, and better manage resources (Smithers, 2004). Versions of national assessments are routinely employed throughout the world, for example, in the U.K., the United States (U.S.), France, Australia, Canada, and Singapore. In England, for example, National Curriculum Tests are applied to all students at the ages of 7, 11, 14, and 16 years in English, mathematics, and science (Jennings, Price, & Pankhurst, 1999).

In the U.S., the National Assessment of Educational Progress (NAEP) collects results from a representative sample across the country (NAEP, 2006). In France, there are diagnostic assessments of all students at the upper secondary levels to identify students' needs and address performance issues (Jennings et al., 1999). Australia has a National Assessment Program, Literacy and Numeracy (NAPLAN) benchmark program, which is administered annually to students in grades 3, 5, 7, and 9 (Australia NAPLAN, 2010). In Canada, the Pan-Canadian Assessment Program (Department of Education) was developed to assess students aged 13 to 15 years in their literacy, science, and mathematics knowledge and skills (Canadian Council of Ministers of Education, 2007). Singapore uses the International Primary School Examination (IPSLE), a nation-wide examination held at the end of primary education to assess pupils' suitability for secondary

education and placement in appropriate courses to match their learning pace, ability, and inclination. Based on their results, candidates are streamed into four different courses: special, express, normal (academic), and normal (technical) (Singaporean Ministry of Education, 2007).

While most of these assessments are not compulsory, other jurisdictions in federal-based countries also have supplementary examinations for local comparisons. For example, in Canada, there are the Ontario District School Board Progress Reports (Ontario Ministry of Education, 2010).

Purpose

National assessments are implemented by administrators and policymakers for a variety of purposes, which Jennings et al. (1999), themed as follows:

- *Formative*: measures students' performances for feedback and guidance toward future learning outcomes.
- *Diagnostic*: to determine students' learning difficulties in order to plan appropriate remedial assistance.
- *Summative*: to systematically record student performance.
- *Evaluative*: to assess measures of schools' performances.

The benefits to be derived from a coherent testing program are proposed by Greaney and Kellaghan (1996, p. 5–10):

1. *Informing policy*: provide factual information to assist policymaking, especially data on the quality of students' learning; to make informed decisions on curricula, resources provisions, and teacher training strategies.
2. *Monitoring standards*: collect regular information on students' performance to assist standards development.
3. *Introducing realistic standards*: foster a sense of realism about appropriate achievement levels. Unrealistic standards contribute to lower performance levels.
4. *Identifying correlation of achievement with outcomes*: correlate a range of factors with outcomes to identify strengths and weaknesses in the system.
5. *Directing teachers and raising students' performance levels*: draw teaching and learning into focus with national assessment to achieve desired outcomes represented in given indicators.

6. *Promoting accountability*: use assessment to justify resource allocation.
7. *Increasing public awareness*: publish national assessments to attract considerable media attention, thus heightening public consciousness on educational matters.
8. *Informing political debate*: direct statistical evidence to issues arising in an educational system—it is more likely to initiate reform.

As a guide, the original aims of the U.S. NAEP in the mid-20th century were both political and technical. Termed “the Nation’s Report Card,” its aims were to track educational reform, educational research, and the war on poverty (Jennings & Renter, 2006). The aims of national assessment in New Zealand’s National Educational Monitoring Project differ in that they are explicit:

...To get a broad picture of the achievements of representative samples of New Zealand school students at successive points in time so that

- trends in educational performance can be identified and reported;
- good information is available to assist policy makers, curriculum specialists and educators with their planning;
- the public can know about trends in educational achievement

(New Zealand Ministry of Education, 2009, webpage).

Capacity to Improve Learning

National assessment may be viewed as both an instrument for reform and a measure of change. In these guises, assessment methodology is open to public scrutiny from a dual perspective regarding purpose and effectiveness (Whetton, Twist, & Sainsbury, 2000). Initially, all national assessments, whether used for formative or summative purposes, are motivated by a desire to improve an educational system (Shepard, 1991). However, a considerable body of research evidence shows that using national assessments for summative purposes (high-stakes tests) are harmful and have negative effects (Elliott & Fuchs, 1997; Frederiksen, 1994; Shepard, 1991). Moreover, teachers, due to the issue of accountability, tend to tailor their instructions to imitate the format of multiple-choice tests (Baker, 1996; Kane, Khattri, Reeve, & Adamson,

1997). They encourage students to focus on what is tested. Therefore, “teaching to the test” is a common practice in schools (Bowers, 1989; Resnick, 1996).

Recognizing the negative effects of standardized exams leads to focus on another proposal that can guide an educational system in a positive direction (Jacobson, 1997; Shepard, 1991). Shepard claims that a system of national examinations has high stakes but avoids the issues created by a narrowed curriculum. The United Kingdom Department of Education and Science (1989) recommends that “the basis of the national assessment system be essentially formative, but designed also to indicate where is a need for more detailed diagnostic assessment” (Jennings et al., 1999, p. 1). Using a national assessment formatively can serve many purposes; encouraging social comment on the direction of a nation’s education focus and administration (Klinger, DeLuca, & Miller, 2008). Educators argue that national assessments will assist in reaching and maintaining higher academic standards than individual establishments are willing to self-impose (Lawton, 1997). Planel, Broadfoot, Osborn, Sharpe, and Ward (2000) argue that a national assessment can be utilized as a case study to demonstrate the fundamental educational and cultural values that underlie an educational system. Assessment results play an essential role in education reform. McKernan (2001) asserted that the educational improvement over the previous decade in the U.S. directly depended on access to high-quality data. None of the accomplishments was possible without sound and objective data to characterize educational progress.

This focus in the U.S. on measurement, planning, and feedback was noted in a National Research Council report, where NAEP models for standards-based reporting were widely adopted among U.S. jurisdictions. States also used these results for curricula and standards development (DeVito & Koenig, 2000). In 1996, the U.K. introduced a national curriculum assessment system based on “attainment targets” in mathematics, English, and science. The overall target for English was that 80 percent of 11-year-olds achieve level 4 or higher by 2002, compared to 63 percent in 1997. By 1999, 70 percent of students reached the target level (Whetton et al., 2000). Performance measurement and implementing progressive targets enabled the U.K. to make considerable progress on TIMSS 2003 scores in mathematics and science. The result of

TIMSS shows that U.K. primary students are among the best performers in mathematics and science, with greater progress in mathematics since TIMSS 1995 than all other contenders, and only two countries scored significantly higher in primary school science. These improvements in performance were achieved throughout the results, for both high and low achievement groups (House of Commons Public Accounts Committee (U.K.), 2009).

Proposed Saudi National Assessment Plan

To address their differing agendas, nations' assessment of student performances as discussed above fit a continuum of purposes, from national aspirations for future prosperity to a set of targets. At this time, however, the Saudi school system retains traditionalist teaching methodology, characterized as follows:

- A highly centralized system, where the Ministry of Education controls all aspects of public education (curriculum, assessment, teachers, facilities, administration, and extraneous funding) and many of the functions of private schools,
- Curricula which lacks rigor, content, and performance standards,
- Inadequate and uncoordinated data which cannot appropriately inform all levels of decision making, and
- Inadequate professionalism in administration of curricula, assessment, and data gathering.

These factors are central to a national assessment strategy for Saudi Arabia. This study offers a proposal which satisfies a strategic view to national assessment, a coherent, comprehensive and phased program supporting a substantive assessment framework. This proposal has a functional aspect which is a national assessment committee; and a policy direction which would take the form of national assessment strategies. The remainder of this study describes both aspects.

National Assessment Committees

The Ministry of Education charged the General Directorate of Evaluation and Educational Quality to develop a plan for national assessment. To provide the resources for developing, implementing, and conduct-

ing national assessments, the following committees are required.

Steering Committee

Comprising influential members internal and external to the Ministry of Education, the role of the steering committee is to facilitate the implementation of national assessments for schools, setting goals and establishing adequate resources. An important function of the steering committee is to provide leadership and assign responsibility to organizations and individuals to address administrative and financial obstacles which can destabilize an assessment program. Further, the committee's role includes transparency, providing a public forum for robust debate of proposals, programs, and outcomes. Initially, the steering committee sets an agenda which includes consideration of these factors in consultation with authorities, professional assessors, and community stakeholders:

- Define the justification and the intent of the national student assessments.
- Define the strategy for content, implementation, and staging.
- Determine terms of reference for national assessment: content, data collection, analysis, and reporting procedures.
- Identify and allocate financial resources and define performance levels.

Implementation Committee

Internationally, student assessments are either implemented by central education authorities, especially in federations where there is a division of control; or administered by authorities, using best practice public or private organizations. In Saudi's case, the Director of Evaluation and Educational Quality is responsible for implementing the national assessments. The implementation group's responsibilities include:

- Project management, consultation with the steering committee and stakeholders, and promoting assessment benefits.
- Identifying facilities and equipment sufficient for national assessment data holding.
- Defining the project, target facilities, or regions for the first stage.
- Curricula analysis, assessment design and approval.

- Determining assessment structures relative to schools, teachers, and activities/questionnaires for students.
- Sampling, making amendments, consultations with the steering committee and stakeholders, and issuing final approvals.
- Providing resources and training for districts, administration, and teachers.
- Developing assessment kits, and producing appropriate and sufficient materials, such as teacher training sets, questionnaires, and activity materials.
- Dissemination, application, and collection of assessment materials.
- Processing and analysis, in consultation with steering committee.
- Reporting of data.

The implementing group comprises subgroups, involving a technical group for instrument development, sampling, psychometrics, and data processing. The technical subgroup, in particular, requires substantial professional input to ensure best practice standards, materials, and training. While more developed countries' educationalists can access professionals locally, and occasionally internally, developing countries seek assistance from external organizations. For example, the U.A.E. contracted the Australian Council for Educational Research (ACER) to assist its Centre for Educational Testing, Measurement and Evaluation (CETME) in the development of assessments for the 2003 year-5 cohort, focusing on English, mathematics, science, and Arabic. This successful program was extended to year-7 U.A.E. students in 2004 and to year-3 and year-9 students in 2006 (ACER, 2007).

However, there is robust debate regarding education authorities undertaking national assessments. Governance is an issue when education authorities audit, or assess, whether self-imposed standards were attained. The bureaucratic processes are not conducive to transparency or adequate consultation with stakeholders. Political influence may occur at any point in the planning or execution of the program (Greaney & Kellaghan, 1996).

National Assessment Strategy

National education systems differ markedly, due to social, economic, and political factors; consequently,

development of national assessment programs vary to meet specific needs. Nevertheless, extant national assessment programs utilize similar frameworks, which is a useful basis for developing a Saudi program respecting Saudi culture and mores. Thus, an assessment program design sponsored by the World Bank (Greaney & Kellaghan, 1996) was the basis for this study's proposal, adapted through reference to subsequent literature, best practice examples from TIMSS, and the specific Saudi environment (see Figure 1, National Assessment Framework for Saudi Schools, page 8).

Assessment Purposes

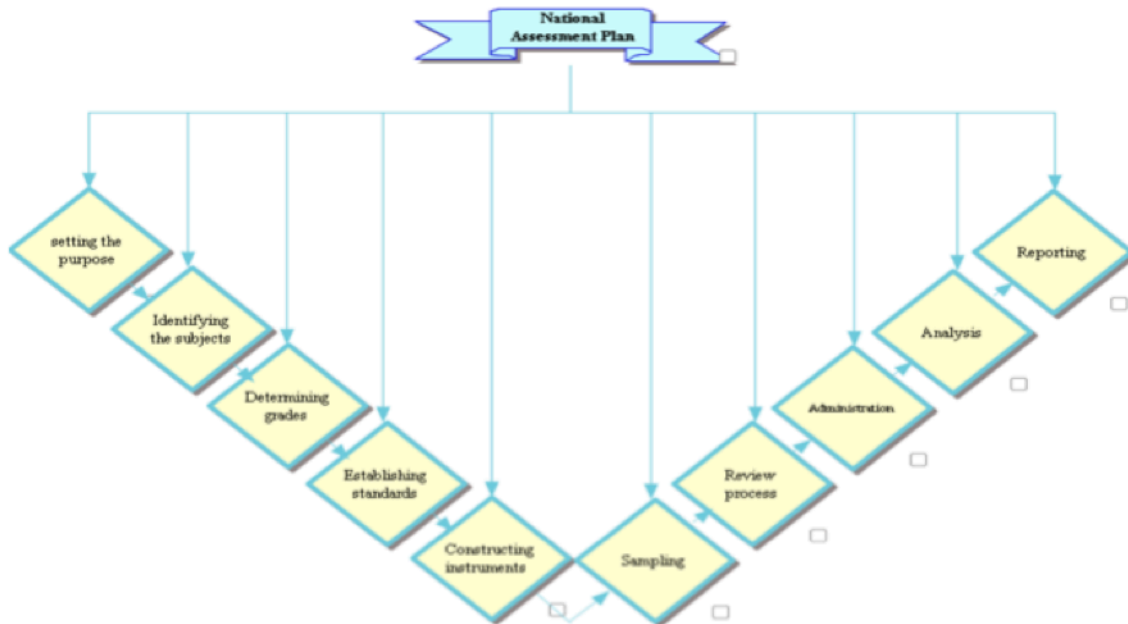
As with all complex undertakings, the goal of national assessments should be simple and manageable. Decision making on possibly a broad range of goals should be influenced by the obstacles ahead, such as the construction of appropriate tests, limited capabilities in computer processing, limited expertise to administer assessment materials, limited resources for data analysis, and the need for consultation with stakeholders. A possible mission statement for the program can be derived from the Ministry of Education's vision: national assessment is a valued contribution to raise Saudi schoolchildren's knowledge and skills to an international competitiveness, especially in science and technology, to successfully make their contribution to the country's future development (Al Hakami, Alkalaf, Ayub, & Alalola, 2006).

Clarity of purpose facilitates the logistics of the program, thus saving time and resources. The objective should be to monitor Saudi students' achievement in key subject areas and to compare that achievement among Saudi schools as well as internationally. Under these objectives, a cascade of priorities emerges; for instance, comparisons can be made between students' achievement in the targeted subjects, between male and female students or between school districts.

Selecting Assessment Subjects

National assessments gather particular grade year information in three distinct areas: 1) cognitive outcomes, specifically subject matter competence; 2) effective outcomes, that is, attitudes toward the subject; and 3) background variables such as school environment. These are discussed in turn.

Figure 1. National Assessment Framework for Saudi Primary Schools



Cognitive Outcomes

Generally, national assessment agencies examine students in their first language (literacy), mathematics (numeracy), and science. The attention to literacy and numeracy reflects the importance of these subjects in the provision of a basic education, and then science emerges as an essential subject in the technology age. The countries that have a long experience in national assessments extend the range of subjects to social studies, or extend the domain of the subject matter (such as, in language, by testing students in speaking, listening, writing, and reading). For example, in Alaska, NAEP conducted assessments in reading, mathematics, science, writing, U.S. history, civics, geography, and the arts (Alaskan Department of Education & Early Development, 2007). In this study, in Saudi education, there are four key subjects for the first stages of national assessment development: mathematics, science, religion, and Arabic.

Effective Outcomes

The majority of national and international assessments (such as TIMSS) measure student attitudes towards subjects, influenced by the assumption that attitudes and interests contribute to students' successful learning. Initially, students' attitudes toward subjects, school, and teachers can be measured; however, measuring the variable of attitudes toward teachers may not give valid information, because the traditional relationship between teacher and student might hinder students' responses.

Background Variables

There are many school and nonschool factors that can be studied in relation to students' achievements, such as the type of school building (governmental or rented), class size, and socioeconomic status.

Selecting Survey Populations

Target populations are usually identified by age, grade level, or both. Choosing a population according to an age factor is disruptive because it requires that students from several grade levels take the test at the same time, and it requires identification of appropriate assessment content for each grade cohort (Greaney & Kellaghan, 1996). In best practice jurisdictions, such as the U.S., the U.K., Australia, and Canada, the grade level is taken into account in defining the population. All countries that conduct national assessments select students in the primary grade years to identify deficiencies at an early point for remedial action. Also, in many countries, national assessments are conducted during the secondary school years, usually at the lower and higher grade levels, to collect indicators regarding students' preparation for life after school. In this study, application of the Saudi assessment program is appropriate for grades 3, 6, 9, and 12, as these years are key stages in public education. However, at the initial stage of assessment development, a focus on the primary level is more useful, proceeding to secondary years' assessment after gaining sufficient skills and experience.

Selecting Assessment Subject Standards

There is increasing pressure on educators to continually raise standards to improve a country's competitive advantage (Whetton, et al., 2000). The formulation of national standards for targeted subjects is essential to establish national assessments. The U.S. Commission on Instructionally Supportive Assessment (2001, p. 3) recommends:

Better teaching and better testing are more likely if educators understand clearly where they need to focus their efforts and test-developers understand the matters they are testing clearly. Therefore, we recommend that states analyze their high-priority content standards to identify what students must do and understand to demonstrate they have achieved the standards.

Further, an indicator of best practice is that primary elements of a national education system are aligned, presenting uniform high standards (National Institute on Educational Governance, Finance, Poli-

cymaking, and Management; Office of Educational Research and Improvement; U.S. Department of Education; and Consortium for Policy Research in Education, 1998). Two types of standards should be established, content standards and performance standards. Content standards govern the parameters of knowledge, understanding, and skills in a specific content area and performance standards provide performance indicators and concrete examples of skills-based learning specified by the content standard (Hawaii Department of Education, n.d.). The U.S. National Research Council and the international TIMSS are two respected sources which explicate standards, and each jurisdiction initiates or adapts their school subject standards from these and other sources. In this study, these sources are useful guides to establishing standards for other subjects such as language and religion.

Constructing Assessment Instruments

Once assessment standards are established, development of instruments begins. A table of specifications, providing essential information regarding the assessment program, facilitates the test development process. This information is then used to focus and guide the remaining steps in the program (Professional Testing, 2006). The test description may simply indicate the content to be covered; that is, the number and percentage of total items on the test address the particular content area at a specific taxonomic level. It can also include elements such as the overall test length, the test administration time limit, and the item types that are expected to be used (for example, multiple-choice, performance assessment). In some cases, the test description may also specify administration mode (for example, paper-and-pencil, performance-based, computer-based), and scoring procedures and scoring rubrics.

Test developers use indicators for levels of cognitive processing that students use in responding to specific items in accord with Bloom's taxonomy; for example, knowledge and application (Krathwohl, 2002). It is critical that the test framework and assessment items include a substantial proportion of items targeted above the knowledge level of cognition. A typical framework is presented as a two-way matrix with the content areas listed in the table rows and the cog-

nitive processes in the table columns. The test blueprint is used to guide and target item writing as well as for test form assembly (Professional Testing, 2006). National assessments rely on two forms: multiple choice, where students are required to select an answer from a list of options; and constructed response, where they provide their own answers (United States National Center for Educational Statistics, 2005). However, constructed-response items require qualified scorers, which are not available at this time in Saudi Arabia.

Sampling

To contain costs in gathering and analyzing data for a national assessment program, and to gain greater speed in data analysis and reporting, a representative sample is used (Greaney & Kellaghan, 1996). For this study, the first step is to select a representative sample of schools, data for which are available from the Ministry of Education. Grades 3, 6, 9, and 12, the key years for schooling, are appropriate targets; the technical committee may differ, or stage the selection over a period of time.

Review for Best Practice

While the program is subject to review by the assessment committees throughout, expert opinion is required of the proposed procedures and assessment instruments to confirm relevance to assessment objectives, adequate benchmarks, and efficiency of operation (Greaney & Kellaghan, 1996). A pilot study precedes the program implementation to ensure that items are clear and understandable, and to fulfill statistical requirements such as validity and reliability. As a guide, all U.S. NAEP items are pilot-tested to 500 responses to evaluate performance before operational use (United States National Center for Educational Statistics, 2005). With TIMSS, the review process included several pilots to ensure that the tests represented the curricula of the participating countries and that the items were timely and did not exhibit any bias towards or against particular countries (Martin et al., 1998; Whetton et al., 2000).

Administration

Substantial logistics are required to conduct a national assessment: gaining agreement with districts and schools to participate; printing, packaging, and distribution of materials; recruiting and training program administrators; supervisory and public relations duties; and processing data (Greaney & Kellaghan, 1996). Involving teachers in conducting tests can reduce administrative costs; it may, however, compromise assessment validity if they do not follow administration procedures adequately. In New Zealand, experienced teachers are selected from a national pool of applicants, attend a week of specialist training, and then work in pairs to conduct the national assessment (New Zealand Ministry of Education, 2008).

Data Analysis

The data processing consists of collecting, scoring, re-recording, and entering data, and establishing a database—followed by analysis. This process requires complex data processing capability and software designed to operate on large quantities of data to mentor the quality of data obtained and to ensure the availability of high-quality data for analysis. For this exacting work, TIMSS employed a rigorous quality control to create its international database, producing manuals and software for data entry from its member countries to gain quality, standardized data for the IEA Data Processing Centre each cycle. Quality controls consist of iterative procedures to identify, document, and correct deviations from the international instruments, file structures, and coding schemes (TIMSS, 2001).

Reporting Results

Assessment results should be reported as soon as possible after data collection. The report should be concise, simply written, and devoid of educational jargon. It should feature simple graphs and bar charts (Greaney & Kellaghan, 1996). The U.S. NAEP reports may serve as models—they include descriptive information about student achievement; evaluative information to support judgments about the adequacy of student performance; and contextual, interpretive information to assist understanding of students' strengths and weaknesses and to gain information relating to policy appli-

cations to address systemic shortfalls (Pellegrino, 2000). For NAEP applications, achievement levels are used to report results in terms of students' knowledge and skills (United States National Center for Educational Statistics, 2006). The conclusion is based on clear evidence derived from the data, and the report should document relevant procedures and criteria.

Conclusion and Recommendations

As a matter of urgency, this study stresses the necessity for Saudi Arabia to implement a national assessment program for primary and secondary school students. Recent TIMSS data exposed a fallacious assumption that the extensive resources and policy commitment placed by the government on its educators has not resulted in adequate returns. Investigation shows that an absence of checks and balances in the Saudi educational system has contributed significantly to this situation, and that a national assessment process comprising agreed-upon standards for school districts, individual schools, and students is a prerequisite to reversing the falling competitiveness of the country's graduates and controlling education expenditures. International organizations such as TIMSS assist implementation of such programs, and best practice expertise is readily available from many developed countries and assessment organizations. Only by implementing a rigorous re-education program, based on assessment, can Saudi youth compete in the international labor markets and manage Saudi Arabia's future prosperity.

References

- Abdaljoad, N. (2005). Primary education (in Arabic). In H. A. Al-Ghamdi & N. Abdaljoad (Eds.), *Development of the Education System in the Kingdom of Saudi Arabia*. Riyadh, KSA: Al-Roashd Library.
- Alaskan Department of Education & Early Development (2007). National Assessments of Education Progress (NAEP). State of Alaska. Retrieved 7 February, 2010 from www.eed.state.ak.us/tls/assessment/naep.html
- Al-Baadi, H. M. (1995). Saudi Arabia. In T. N. Postlethwaite (Ed.), *International Encyclopedia of National Systems of Education* (2nd ed.) Oxford, England: Pergamon.
- Al-Daoud, H. (2004). *The reality of an ongoing assessment of mathematics in the early grades of primary education for girls*. Unpublished master's thesis, King Saoud University, Riyadh, KSA.
- Al-Dossary, A. (2000). *Handbook for educational evaluation* (3rd ed.). Riyadh, KSA: Arabic Educational Office for the Gulf States.
- Al-Ebarahim, H. A. M. (2001). *The extent of the effectiveness of the student evaluation register in improving achievement levels in the general education phases for girls*. Unpublished master's thesis, King Saoud University, Riyadh, KSA.
- Al-Ghamdi, H. A. (2005). Educational policy in the Kingdom of Saudi Arabia (in Arabic). In H. A. Al-Ghamdi & N. Abdaljoad (Eds.), *Development of the Education System in the Kingdom of Saudi Arabia*. Riyadh, KSA: Al-Roashd Library.
- Alhakami, A. S. (2004). *Educational evaluation in the Kingdom of Saudi Arabia: the study of reality and the vision of the future strategy*. Paper presented at the Evaluation Strategies to Achieve Comprehensive Quality in Education program, Beirut, Lebanon.
- Alhakhmi, A., Alkalaf, A., Ayub, H., & Alalola, M. (2006) *National tests to measure students' achievement*. Riyadh, KSA: King Abdulaziz City for Science and Technology.
- Al Hakmi, A. B. S. (2000). *National report: Kingdom of Saudi Arabia, "Education for all" assessment for the year 2000*. UNESCO. Retrieved February 5, 2010, from www.unesco.org/education/wef/countryreports/saudi_arabia/contents.html
- Al-Hamed, M. M. (2005). The educational innovations in Saudi Education (in Arabic). In M. M. Al-Hamed, M. P. Zeada, B. G. Al-Otaibi, & N. A. Metwalli (Eds.), *Education in the Kingdom of Saudi Arabia: The vision of the present and exploration of the future*. Riyadh, KSA: Al-Roashd Library.
- Almegidi, A. M. (2004). *The development of the educational system in Saudi Arabia*. Riyadh, KSA Al-Roashd Library.
- Al-Otaibi, B. G. (2005). Problems of education in the Kingdom of Saudi Arabia (in Arabic). In M. M. Al-Hamed, M. P. Zeada, B. G. Al-Otaibi & N. A. Metwalli (Eds.), *Education in the Kingdom of Saudi Arabia: The vision of the present and exploring the future*. Riyadh, KSA: Al-Roashd Library.

- Al-Sadan, I. A. (2000). Educational assessment in Saudi Arabian schools. *Assessment in Education: Principles, Policy & Practice*, 7(1), 143–155.
- Australia National Assessment Program, Literacy and Numeracy (AAPLAN). (2010). *NAPLAN 2010 tests*. Retrieved 6 February, 2010 from www.naplan.edu.au/.
- Australian Council for Educational Research (ACER). (2007). *United Arab Emirates national assessment of student achievement and progress*. Retrieved February 7, 2010, from www.acer.edu.au/enews/0407_UAE_MOU.html.
- Bahgat, G. (1999). Education in the Gulf monarchies: retrospect and prospect. *International Review of Education*, 45(2), 127–136.
- Baker, E. (1996). What the research says about student assessment. *Improving America's Schools: A Newsletter on Issues in School Reform*. Retrieved February 7, 2010, from www2.ed.gov/pubs/IASA/newsletters/assess/pt4.html
- Bowers, B. C. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41–47.
- Canadian Council of Ministers of Education. (2007). *National and international assessments*. Retrieved February 6, 2010 from www.cmec.ca/Programs/assessment/pancan/Pages/default.aspx
- DeVito, P. J. & Koenig, J. A. (2000). *Designing a market basket for NAEP: Summary of a workshop*. Washington, DC: National Academies Press.
- Elliott, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review*, 26(2), 224–233.
- Fensham, P. (1998). Student response to the TIMSS test. *Research in Science Education*, 28(4), 481–489.
- Frederiksen, N. (1994). *The influence of minimum competency tests on teaching and learning*. Princeton, NJ.: Educational Testing Service, Policy Information Center.
- Greaney, V., & Kellaghan, T. (1996). *Monitoring the learning outcomes of education systems*. Washington, DC: The World Bank.
- Hawaii Department of Education (n.d.). *Hawaii's implementation of standards-based education*. Retrieved February 7, 2010, from <http://doe.k12.hi.us/curriculum>
- House of Commons Public Accounts Committee (U.K.), (2009). *Mathematics performance in primary schools: Getting the best results*. London, England: Author.
- Jacobson, L. (1997). Principals' survey finds support for standards, voluntary national tests. *Education Week*, 16(30), 7.
- Jakwerth, P. (1999). TIMSS performance assessment results: United States. *Studies in Educational Evaluation*, 25(3), 277–281.
- Jennings, S., Price, N., & Pankhurst, K. (1999). To what extent can national curriculum tests in mathematics inform and guide teaching? *International Journal of Mathematical Education in Science and Technology*, 30(1), 1–10.
- Jennings, J. & Rentner, D. (2006). Ten big effects of the No Child Left Behind Act on public schools. *Phi Delta Kappan*, 88(2), 110–113.
- Kane, M., Khattri, N., Reeve, A., & Adamson, R. (1997). *Assessment of student performance: studies of education reform*. Washington, DC: Office of Educational Research and Improvement, US Department of Education.
- Kind, P. M. (1999). Performance assessment in science—What are we measuring? *Studies in Educational Evaluation*, 25(3), 179–194.
- Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, 76. Retrieved 2 April, 2010 <http://www.umanitoba.ca/publications/cjeap/articles/klinger.html>.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice*, 41(4), 212–218.
- Lawton, M. (1997). Experts question value of new national tests. *Education Week*, 17(1), 1.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1998). *Science achievement in the primary school years: IEA's third international mathematics and science report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

- Martin, M. O., Mullis, I. V. S., & Foy, P. (2009). *TIMSS 2007 international science report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades* (Revised). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international science report. Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- McKernan, G. J. (2001). *Report of the measuring success task force to the national education goals panel*. Not retrievable as Panel is defunct: <http://govinfo.library.unt.edu/negp/page9-3.htm>
- Ministry of Education. (2001). *Education in the Kingdom of Saudi Arabia within the last hundred years*. Riyadh, KSA: Farazdk Press.
- Ministry of Education. (2004). *The development of education: national report of the Kingdom of Saudi Arabia (2000-2004)*. Riyadh, KSA: Dar Al-Hilal Press.
- Ministry of Education. (2005). *Fifty years of education*. Riyadh, KSA: Dar Al-Hilal Press.
- Ministry of Education. (2006). *A statistical summary report on the education of boys and girls in Saudi. Riyadh*. KSA: Dar Al-Hilal Press.
- Ministry of Economy and Planning (2005). Preliminary results of the census operation 1425 H-2005. Retrieved April 14,2010, from <http://www.cdsi.gov.sa/showproductstandard.aspx?id=26&pid=1650>
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2009). *TIMSS 2007 international mathematics report: findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. (Revised). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- National Assessment of Educational Progress (NAEP). (2006). *State profiles*. Retrieved February 6, 2010, from <http://nces.ed.gov/nationsreportcard/about/statepartic.asp>
- National Institute on Educational Governance, Finance, Policymaking, and Management; Office of Educational Research and Improvement ;U.S. Department of Education; and Consortium for Policy Research in Education (1998). *Policy Brief: What the Third International Mathematics and Science Study (TIMSS) Means for Systemic School Improvement*. Retrieved April 14,2010, from <http://www2.ed.gov/PDFDocs/timssbrief.pdf>
- New Zealand Ministry of Education. (2008). *The national education monitoring project. Projects update*. Retrieved February 7, 2010, from www.educationcounts.govt.nz/publications/assessment/latest_nemp_reports
- New Zealand Ministry of Education, (2009). *About NEMP: an overview*. Retrieved February 7, 2010 from http://nemp.otago.ac.nz/_about.htm.
- Ontario Ministry of Education. (2010). *School board progress reports*. Retrieved February 6, 2010 from www.edu.gov.on.ca/eng/bpr/.
- Pellegrino, J. W. (2000). *Grading the nation's report card: evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academies Press.
- Panel, C., Broadfoot, P., Osborn, M., Sharpe, K., & Ward, B. (2000). National assessments: underlying cultural values revealed by comparing English and French national tests. *European Journal of Education, 35*(3), 361–374.
- Professional Testing (2006). *Step 3: Create the test specifications*. Retrieved February 7, 2010, from www.proftesting.com/test_topics/pdfs/steps_3.pdf
- Resnick, M. (1996). Towards a practice of "constructive design". In L. Schauble & R. Glaser (Eds.), *Innovations in Learning: New Environments in Education*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rugh, W.A. (2002). Education in Saudi Arabia: choices and constraints. *Middle East Policy, 9*(2), 40.
- Shepard, L. A. (1991). Will national tests improve student learning? *Phi Delta Kappan, 73*(3), 232–238.

- Siddiqui, M. A. (1996). Library and information science education in Saudi Arabia. *Education for Information, 14*(3), 195–214.
- Singaporean Ministry of Education. (2007). *Singapore examinations and assessment board. International Primary School Leaving Examination*. Retrieved February 7, 2010, from www.seab.gov.sg/SEAB/ipsle/ipsle.html
- Smithers, A. (2004). Secondary education: teacher retention and recruitment. *House of Commons Education and Skills Committee Fifth Report of Session 2003-04*. (Vol. II, HC 1057-11, Ev 230-237). London, England: The Stationery Office.
- Trends in International Mathematics and Science Study (TIMSS). (2001). *TIMSS 1999 science benchmarking report*. Retrieved February 7, 2010, from http://timss.bc.edu/timss1999b/sciencebenchmarking/t99bscience_TOC.html
- United Kingdom Department of Education and Science (1989). *National Curriculum Task Group on Assessment and Testing: A report*. London, England: Author.
- United Nations Development Program (UNDP). (2005). *Program on governance in the Arab region*. Retrieved February 7, 2010, from www.undp-pogar.org/countries/indicator.asp?ind31&tyr2005#notes#notes
- United Nations Development Program. (2009). *Human development report*. Retrieved February 6, 2010, from <http://hdrstats.undp.org/en/indicators/99.html>
- United States–Saudi Business Council. (2010) *Saudi Arabia releases 2010 budget*. Retrieved February 6, 2010, from www.us-sabc.org/custom/news/details.cfm?id559
- United States Commission on Instructionally Supportive Assessment. (2001). *Building testing to support instruction and accountability: A guide for policymakers and "illustrative language for an RFP to build tests to support instruction and accountability."* Retrieved February 7, 2010, from www.testaccountability.org/
- United States National Center for Educational Statistics (2005). *NAEP item scoring process*. Retrieved February 7, 2010, from http://nces.ed.gov/nationsreportcard/contracts/item_score.asp
- United States National Center for Educational Statistics (2006). *NAEP frameworks*. Retrieved February 7, 2010, from <http://nces.ed.gov/nationsreportcard/frameworks.asp>
- Whetton, C., Twist, E., & Sainsbury, M. (2000, 24–28 April). *National tests and target setting: maintaining consistent standards*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.