

USING THE RESULTS OF A NATIONAL ASSESSMENT OF EDUCATIONAL ACHIEVEMENT

**Thomas Kellaghan
Vincent Greaney
T. Scott Murray**

Chapter 1

Factors affecting the use and nonuse of national assessment findings

The main objectives of a national assessment, as set out in volume 1 of this series, *Assessing National Achievement Levels in Education*, are to determine (a) how well students are learning in the education system (with reference to general expectations, aims of the curriculum, and preparation for further learning and for life); (b) whether there is evidence of particular strengths and weaknesses in students' knowledge and skills; (c) whether particular subgroups in the population perform poorly; (d) which factors are associated with student achievement; (e) whether government standards are being met in the provision of resources; and (f) whether the achievements of students change over time (Greaney and Kellaghan 2008). In pursuit of these objectives, through procedures established in the social sciences, data are collected from students and other stakeholders in the education system. Such data collection serves to make the outcomes of educational management and practice more transparent and has the ultimate purpose of providing personnel in the system with information designed to improve their practice (Ferrer 2006). Evidence on the attainment of the objectives of a national assessment has implications for assessing important aspects of how an education system functions with respect to access, quality, efficiency, and equity (Braun and others 2006) (see box 1.1). The assessment will more than likely find that the issues are interrelated. In many education systems, low-achieving schools tend to serve students from disadvantaged backgrounds or a minority group; to receive the lowest level of resources (for example, textbooks may arrive late, if at all); and to have difficulty attracting teachers because of isolated location or for ethnic or language reasons. Clearly, any information that a national assessment can provide about these issues should be of interest to a wide range of stakeholders: politicians, education managers, teachers, teacher trainers, curriculum developers, parents, employers, and the general public.

Earlier books in this series described how information is obtained in a national assessment: how instruments to collect information on student achievement and associated variables are designed; how a sample of students is selected to represent the achievements of the education system as a whole (or a clearly defined part of it, such as grade 4 students or 11-year-olds); what procedures should be followed in collecting and cleaning data; and what methods may be used to analyze the data. This book turns to the reporting and use of data obtained in a national assessment with the ultimate objective of improving the quality of students' learning. It is intended for two primary readerships: (a) those who have responsibility for preparing assessment reports and for communicating and disseminating findings and (b) users of those findings. This introductory chapter addresses five topics. First, it describes aspects of the political context in which a national assessment is carried out and their implications for using assessment findings. Second, it discusses the issue of accountability, which is a major concern in many government administrations and one with which national assessment activities have been closely identified. Third, it notes that the quality of the instruments used in a national assessment to obtain information related to students' learning (the knowledge, skills, attitudes, and habits that students have acquired as a result of their schooling) has important implications for the use of findings to improve learning. Fourth, it considers how characteristics of a national assessment (census based, sample based, or international) affect the way findings can be used. Finally, it outlines possible reasons for the lack of use of national assessment findings.

BOX 1.1

Issues on Which a National Assessment Can Throw Light

Access. Obstacles to attending school, such as limited availability of places or distance of students' homes from school.

Quality. The quality of inputs to and outputs of schooling, such as the resources and facilities available to support learning (responsive curricula, teacher competence, textbooks); instructional practices; learner-teacher interactions; and student learning.

Efficiency. Optimal use of human and financial resources, reflected in pupil-teacher ratio, throughput of students, and grade repetition rates.

Equity. Provision of educational opportunities to students and attainment of parity of achievement for students, irrespective of their characteristics, such as gender, language or ethnic group membership, and geographic location. Source: Authors based on Braun and others 2006.

The Political Context of a National Assessment

Although one national assessment may look very much like another in many respects, there are, in fact, differences between assessments that have implications

for use. Differences in design, implementation, and use arise from the fact that assessment is a political phenomenon (as well as a technical one), reflecting the agenda, tensions, institutional norms, and nature of power relations between political actors. Identifying the political context in which an assessment is carried out can help explain differences between countries in their evaluation strategies (Benveniste 2002). Even within the United States, accountability systems differ from state to state, reflecting administrative decisions and traditions that have evolved over time (Linn 2005b). The role of assessment (and evaluation) in the exercise of control and power in educational matters has several facets. In the first place, assessment originates in a political process, often inspired and fashioned by political motivations. Second, the form of an assessment will be the result of competition among social actors who vie to influence the determination of norms and values that the state will privilege. Third, an assessment can affect social relations between, for example, education managers and teachers or teachers and parents. Fourth, control over the disposition and interpretation of assessment outcomes signifies authority to influence policy, resource allocation, and public perceptions. Finally, an assessment can involve mechanisms for regulation and for holding social actors accountable, implicitly or explicitly, for outcomes (Benveniste 2002). The social actors with the ability to influence the nature of an assessment and the ways findings are used are many. How power politics actually play out in a country will depend on a number of factors, such as the following:

- The extent to which decisions regarding educational provision (for example, financing, curricula) are the function of central or decentralized governance
- The existence and strength of informal institutions, networks, and special interest groups, both within and outside government
- The strength of teachers' unions, which can play a key role in policy implementation, if not in policy formation
- The role of external (multilateral and bilateral) agencies in sensitizing administrations to address conditions in their education system and in providing or supporting development of the capacity to deal with them. The implications of a political stance for a national assessment can be illustrated by two examples (Benveniste 2002). In Chile, emphasis is placed on accountability to the public, thereby promoting market competition, which is induced by publication of the results of an assessment for individual schools. Quite a different stance is represented in Uruguay, in which the state accepts responsibility for student achievement and for providing the resources required to support student learning--particularly that of the most underprivileged sectors of the population. A further important aspect of the political context in which a national assessment is carried out that has implications for the use of findings is the extent to which an education system is open or closed. Some systems have been described as "exclusionary." In such systems, access to important information about aspects of the education system, including results of research, is limited to policy elites or senior decision makers, who do not permit public dissemination. At the other extreme, in more open systems, efforts will be made

to attract the interest of the media, to mobilize political forces, and to generate debate about educational matters (Reimers 2003). An intermediate position relevant to the use of national assessment data is one in which the circulation of information about the education system, including student achievement data, while not totally restricted, is limited. For example, in Uruguay, student achievement data are in-tended primarily for consumption within the education community (Benveniste 2002).

Accountability

Accountability movements, in response to political, social, and economic pressures, have in recent decades attained increasing importance in government administrations in many countries. This section considers accountability in the context of education and, in particular, focuses on the ways interpretation of the concept affects the use of national assessment data. It should be borne in mind that much of the discourse is based on experience in the United States and focuses on the accountability of schools (McDonnell 2005). The development of accountability movements in the public sector (including education) can be related to a variety of factors that are not mutually exclusive, including the following:

- The need to manage finite (and in some cases decreasing) resources and to increase output for a given amount of input.
- The use of planning and management ideas that are borrowed from the business world, particularly ones relating to quality assurance, customer satisfaction, and continuous improvement (features of the New Public Management movement and a corporatist approach to administration). Such concepts may, in turn, involve defining performance in terms of results, setting performance targets, using performance indicators to determine the extent to which targets are met, implementing strategic and operational planning, and basing resource allocation on performance.
- The introduction of market mechanisms of distribution and control involving incentive schemes, competition, contracting, and auditing, and the transfer of power relations into self-control mechanisms in an effort to minimize the need for external surveillance and to make individuals internalize the norms, values, and expectations of stake holders and the mentality required to govern themselves.
- A movement toward more evidence-based practice. Such a movement requires data to support claims that individuals or institutions have performed professionally and efficiently, as well as data on which to base decisions regarding resource allocation (see Blalock 1999; Clegg and Clarke 2001; Davies 1999; Hopmann and Brinek 2007; Kellaghan and Madaus 2000). A national assessment fits well with many of these factors by providing relatively simple statistical information (evidence) about the education system on a timely basis. Furthermore, it can identify subgroups or units in the population that meet a specified standard and ones that do not. The information can be used for planning and management—in particular, for deciding on the action required to improve quality or efficiency. It can also be used to hold social actors implicitly or explicitly accountable, thus

placing on them the onus for change or adjustment. The focus on accountability varies throughout the world, unfolding with different speeds and impact (Hopmann and Brinek 2007). It is thus not surprising that the purposes and goals of many national assessments—particularly in developing countries or the ways such assessments fit into a system of accountability may not be very clear. In a situation in which accountability policies are not well developed, national assessment findings are unlikely to have much effect (Hopmann and Brinek 2007). However, at least an implicit recognition of accountability would seem to be necessary if use is to be made of assessment results. Otherwise, how are decisions to be made about the action that needs to be taken following an assessment and the individuals or institutions that will take the action?

Assigning accountability to the many stakeholders involved in a system as complex as education is not a trivial matter. Six questions that can help clarify the issues involved merit consideration in this task particularly when national assessment results are used to hold schools and teachers accountable.

Should an Accountability System Focus on Outcomes?

A focus on the outcomes of education in particular, student learning can be attributed to the formal recognition and concern that many children spend a considerable amount of time in school without acquiring useful knowledge and skills. The need to ensure that children actually learn as a result of their educational experiences was highlighted at the World Conference on Education for All held in Jomtien, Thailand, in 1990 (UNESCO 1990) and again in the Dakar Framework for Action (UNESCO 2000). To use data on the outcomes of education as the sole basis of accountability, however, is to lose sight of the fact that aspects of provision (for example, school buildings, curricula, educational materials, teachers' instructional techniques, and preparation activities) are also relevant in assessing quality. These factors are important if for no other reason than that the quality of student learning depends on them. Students “cannot be expected to become proficient unless and until the content and process of their classroom instruction well prepares them to do so” (Haertel and Herman 2005: 21).

Should an Accountability System Focus on Cognitive Outcomes?

Most people would probably agree that schooling has many purposes – some personal (for example, students' cognitive, moral, and social development) and some societal (for example, promoting social cohesion or nation building). Most would probably also agree that cognitive outcomes are preeminent and, moreover, that development of the literacy and numeracy skills measured in all national assessments is necessary as a foundation for students' later educational progress. It could hardly be considered satisfactory, however, if total reliance on these measures for accountability purposes were to result in the neglect of other valued outcomes of schooling related to attitudes, values, motivation, aspirations, self-concept, ability to work in groups, oral presentation skills, and socialization. Employers and

economists have identified many of these outcomes (often described as soft skills) as very important in gaining employment (Cheng and Yip 2006).

Should an Accountability System Be Based on a Single Measure of Student Achievement?

In most national assessments, a single test (though it may have a number of forms) is used to assess students' competence in a curriculum area (for example, mathematics, reading, or science). Thus, a question arises: even if students' cognitive achievement is accepted as a legitimate criterion of the quality of schooling, is it reasonable to base the assessment of that quality (and a possible assigning of accountability) on a single measure of the performance of students at one or two grade levels? The answer would seem to be no. A test can provide only a limited amount of information about student achievement (see box 1.2). An inaccurate picture of student learning, whether learning is being assessed at the national level or at the level of the individual school, requires multiple measures of achievement (Guilfoyle 2006). If a test is limited to multiple-choice items, additional problems are likely to arise, because it is extremely difficult, using that format, to measure higher-level cognitive skills.

BOX 1.2

Evaluating Achievement on Limited Evidence "Any system that hinges the evaluation of an entire school on one test score average from one group of students at one grade level cannot hope to accurately assess that school."

Source: Guilfoyle 2006: 13.

Should Sanctions Be Attached to Performance on a National Assessment?

A key decision in the use of national assessment findings is whether sanctions should be attached to student performance. Although some attribution of accountability, even if not explicitly acknowledged, might be expected after an assessment, it does not necessarily follow that sanctions will be applied. In some national assessments, however, sanctions are applied, usually to schools, teachers, and in some cases students. Examples of such instances can be found in the assessment of the national curriculum in England, which was introduced primarily as a tool of accountability, and in several state-level assessments in the United States. In such cases, an assessment becomes a high-stakes operation for schools, with a variety of rewards or punishments attached to student performance. Schools or teachers may receive rewards in the form of monetary bonuses, teachers may be dismissed, and students may be denied promotion or graduation. A number of arguments support the attachment of high stakes to student performance on a test. First, it encourages individuals (in particular, teachers) to internalize the norms, values, and expectations of stakeholders (in particular, those of the ministry of education) and to accept responsibility for conforming to them. Second, it supports the operation of market mechanisms in the education system, involving competition, contracting,

and auditing. Third, it serves to focus teacher and student endeavors on the goals of instruction and to provide standards of expected achievement that students and teachers can aspire to, thus creating a system of measurement-driven instruction. In this situation, one might reasonably expect student performance to improve if instruction has been closely aligned with an assessment instrument. Improved performance, however, may not be evident when student achievement is assessed on other instruments. When the achievement gains over time on the U.S. National Assessment of Educational Progress of students in states that have high stakes attached to their state-level assessments are compared with the gains of students in states that do not have high-stakes testing, the findings are ambiguous (Amrein and Berliner 2002; Braun 2004). Arguments against attaching high stakes to students' test performance are based, for the most part, on observation and research on public examinations (rather than on national assessments) over a long period of time (Kellaghan and Greaney 1992; Madaus and Kellaghan 1992; Madaus, Russell, and Higgins 2009). Similar conclusions are emerging about the effects of testing associated with the No Child Left Behind legislation in the United States (Guilfoyle 2006). The available evidence indicates that when sanctions are attached to student performance, negative consequences follow:

- Teachers will tend to react by aligning their teaching to the knowledge and skills assessed in the test (“teaching to the test”), thus neglecting curriculum areas (for example, art, social studies, physical education) that are not assessed.

- Teaching will tend to emphasize rote memorization, routine drilling, and accumulation of factual knowledge, resulting in a passive approach to learning, rather than an approach that stresses higherorder general reasoning and problem-solving skills.

- Teachers are likely to spend considerable time developing students' test-taking strategies (such as how to answer multiple-choice questions) and may even use the multiple-choice format in their teaching (see box 1.3).

BOX 1.3

The Negative Effect of High-Stakes Accountability Uses

“Assessment systems that are useful monitors lose much of their dependability and credibility for that purpose when high stakes are attached to them. The unintended negative effects of the high-stakes accountability uses often outweigh the intended positive effects.”

Source: Linn 2000: 14.

Should League Tables Be Published Following a National Assessment?

A particular example of the use of high stakes in a national assessment is the publication of results in the form of league tables in which schools are ranked in the order of their performance. The expectation of this approach is that it will induce competition among schools and, in turn, improve student achievement (Reimers 2003). The information can

be used to inform parents and communities, and in some situations, parents can use the information to make choices about schools for their children. Even when school choice is not an option or when parents do not use assessment results to make such a choice (Vegas and Petrow 2008), the mere publication of information about the performance of schools can pressure schools to improve their performance. In addition to the adverse impacts on teaching and learning that have already been listed in relation to high-stakes assessment procedures, several other problems can be anticipated when results are calculated and published for individual schools (Clotfelter and Ladd 1996; Kane and Staiger 2002; Kellaghan and Greaney 2001; Linn 2000). First, the performance of schools (and thus their position in a league table) may vary depending on the outcome that is assessed (for example, reading or mathematics achievement). Second, even rankings that are based on the same measure can vary depending on the criterion of “success” that is used (for example, mean score or the proportion of students who obtain “high” scores). Third, the lack of precision in assessment procedures means that small differences between schools (which can have a large impact on their rank) will be due to chance. Fourth, achievement scores can vary from year to year because of factors that are outside the control of the school (for example, differences in cohorts of students). Small schools are particularly vulnerable to this problem. Fifth, the achievements of students in a school represent more than the efforts of teachers, as illustrated by the fact that school rankings based on achievement and socioeconomic data are almost identical (Vegas and Petrow 2008). Sixth, to take account of factors over which the school has no control (for example, student ability, home environment), the mean of gains in student test scores during a year may be used as the index of a school’s performance. However, this measure tends to show very little between-school variance and has been therefore found unsatisfactory. Furthermore, it does not take into account the fact that the rate of students’ growth is related to their initial achievements. More sophisticated statistical approaches, which take into account a range of factors over which schools do not have control, may be used in calculating school gain scores (value-added models). Problems that arise with these approaches are the complexity of the administrative procedures needed to collect the data, the level of statistical expertise required for analysis, the difficulties in the choice of variables to be included in statistical models, and the fact that adjustment for prior achievement may result in lower expectations for low-achieving students. Finally, league tables invite corrupt practices, such as ensuring that low-achieving students do not take part in the assessment or focusing on the performance of borderline students to boost the mean score of a school. False information on conditions in schools may be provided (as occurred in Chile) to manipulate the socioeconomic categorization of the school if a low category attracts benefits.

Who Should Be Regarded as Accountable?

A major argument against attaching high stakes for schools and teachers to student performance in a national assessment is that an assessment does not identify the aspects

of achievement that can be attributed to schools or teachers. Even a cursory reflection on the wide range of factors that interact to affect student performance should cause one to pause before assigning accountability. The factors can be identified as (a) characteristics of students, including their earlier achievements; (b) conditions in which students live, including family and community resources and support; (c) education policies and the resources and support, including curricula and teacher preparation, that are provided by the relevant public authorities; (d) school conditions and resources, including governance and management; and (e) competence of teachers (Kellaghan and Greaney 2001). It seems reasonable to expect that the individuals or institutions associated with these factors should be held responsible and accountable only for the matters over which they have control. Thus, responsibility is shared by (a) students; (b) teachers; (c) schools; (d) policy makers, administrators, and managers of the school system (at national, state, regional, or municipal level, depending on how the education system is organized); (e) providers of support services (curriculum developers, teacher trainers, and textbook publishers); (f) parents; and (g) others (including politicians, taxpayers, and the society at large). In fact, it is extremely difficult to apportion accountability among this variety of stakeholders (see box 1.4). Failure to recognize this problem, however, may lead to incorrect attribution, which, in turn, may result in inappropriate action (see box 1.5).

BOX 1.4

Assigning Accountability for Student Achievement

Assigning accountability for student achievement is not easy:

– Although teachers are accountable to some degree for the achievements of their students, does this mean that they should be held solely accountable if the school in which they teach lacks basic amenities? If they have been inadequately trained? If students are absent for long periods? If the curriculum that they are required to teach is unsuitable for their students?

– Should students be held accountable if their homes do not value education or provide facilities or opportunities to study? If the curriculum is unsuitable? If their early educational experiences were marred by incompetent teaching?

– Should parents be held accountable if they do not send their children to school regularly because they cannot afford the costs involved or because they need the children to work?

– Should education managers be held accountable if they do not provide sufficient funds to meet the needs of schools?

– Should politicians be held accountable if the money available to government is inadequate to meet the demands of the education system?

Source: Authors.

BOX 1.5

Shifting the Blame for School Failure

"Accountability testing may subtly shift the blame for school failure from inadequate school resources, poor teacher preparation, or out-of-school factors to teachers and students who are 'simply not working hard enough,' and thereby divert attention from more costly, more needed reforms."

Source: Haertel and Herman 2005: 3.

Many national assessments, at least implicitly, recognize the role of factors outside the school in determining student achievement. Even in high-stakes assessments, test results are often presented separately for schools, depending on the socioeconomic status of the students whom they serve. Students' scores may also be adjusted to take account of the characteristics of students, such as prior achievements or the socioeconomic status of their families. Moreover, even when high stakes are attached, additional resources will usually be provided in schools experiencing difficulty. Such action accepts that teachers who are not performing well may need assistance and sustained professional development (Linn 2000). A consideration of the variety of stakeholders that can affect students' learning supports the conclusion that assessing accountability is a complex matter and should not be based on the limited statistics that a national assessment provides. In the case of teachers, assessing accountability requires a clinical judgment that takes account of a range of factors, including the circumstances in which they teach. Such judgments are best made by a professional (a head teacher, inspector, or supervisor). Deciding on the accountability of other stakeholders is equally complex. Politicians are accountable to the electorate in a democratic system, but it is far from obvious what priority citizens give to education, much less to the achievements of students, when they cast their votes. Education managers are accountable to their superiors and political masters for the performance of their duties, but again it is not obvious whether student achievement should be a consideration. The remoteness of education managers from the actual work of the school, in contrast to the position of teachers, would probably ensure that student achievement does not play a role in assessing their performance. Greater clarity and transparency about the responsibility and accountability of all individuals and institutions that contribute to the outcomes of the education system (including student learning) should serve to remove many of the ambiguities that exist in current accountability systems. Furthermore, use of an accountability system that includes all individuals, institutions, and agencies that exercise control over the resources and activities of schools should serve to focus the energies of all involved on performing the tasks for which they are responsible (Clegg and Clarke 2001).

The Quality of the Assessment Instrument

The term quality applies to a variety of aspects of students' educational experiences, including learning environments that are safe and adequately resourced, curricula that are

responsive to students' needs, instructional practices, competent teachers who engage in active pedagogies, and students' learning (see, for example, Schubert 2005; UNESCO 2000; UNICEF 2000). In national assessment studies, however, as we have seen, the major focus when considering quality is on cognitive outcomes of the educational process—that is, what students have learned with a view to developing strategies to improve those outcomes. This emphasis is in keeping with target 6 of the Dakar Framework for Action, which highlights improving the quality of education “so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy, and essential life skills” (UNESCO 2000: 8). In recognition of the central role accorded to student learning in a national assessment, this section describes four conditions that should be met to ensure (a) that the test that is used accurately represents the achievements that schools strive to develop and (b) that the information obtained serves the needs of users (Beaton and Johnson 1992). First, because a test can measure only part of the knowledge and skills specified in a curriculum or a construct (for example, reading), ensuring that it provides an adequate representation of that knowledge and those skills is important (see Haertel and Herman 2005; Linn and Baker 1996; Messick 1989). Furthermore, test items should exhibit curricular importance, cognitive complexity, linguistic appropriateness, and meaningfulness for students. Hence, a test should not be limited to measuring isolated skill components or items of knowledge that require students only to recall facts or information (a feature of many national assessments) if the goal of the education system is to develop higher-level cognitive skills (involving reasoning, the ability to identify and solve problems, and the ability to perform nonroutine tasks). Test developers should have in mind the desirability of devising an instrument that will provide a basis for policy and decisions that are likely to induce curriculum and instructional changes that, in turn, foster the development of valued knowledge and skills (see Frederiksen and Collins 1989). To secure adequate representation of a domain or construct or of objectives or subdomains (for example, content strands or skills in mathematics) – a test should contain an adequate number of items. The small number of items in some national assessments must raise questions about their adequacy in this respect. For example, the number of items in tests in Latin American assessments (2040, except in Brazil) means that content coverage has been poor. Furthermore, it is difficult to justify a view that mastery of a specific objective can be determined with only three or four items (González 2002). This sort of inadequacy is by no means limited to national assessments in Latin America. Second, a test should assess knowledge and skills at a level that is appropriate for the students who will take it. A problem will arise if a test is based solely on curriculum documents if the curriculum contains unrealistic expectations for student achievement. In this situation, which is fairly common in developing countries, the test will be much too difficult for lower-achieving students and will fail to register their accomplishments. The solution lies in taking into account in test development not just the standards of the intended curriculum, but also what is known of the actual achievements of students in schools. In practical terms, very small proportions of students should get all the items right or all the items wrong. This

result can be achieved by involving practicing teachers in the development and selection of test items and by carefully field-trialing items before the main assessment in a sample of students that spans the variation in schools of the target population. A third condition that should be met if one is to be confident that a test provides valid information on students' knowledge and skills in a particular curriculum domain is that students' performance should not be determined by their competence in domains other than the one that the test was designed to assess (Messick 1989). For example, a test designed to assess students' achievement in science or mathematics should not contain so much language that performance on it depends on the differential ability of students to read rather than on their ability in science or mathematics. This problem occurs when it cannot be assumed that all students responding to the test possess the same level of skill in reading, which probably would be the case when the language of the test differs for some students from that which they normally use. Finally, if assessment results are to be used to monitor change over time, the assessment instruments must be comparable. To achieve this result, the same test, which should be kept secure between administrations, may be used. If different tests are used, scaling with Item Response Theory allows results to be presented on the same proficiency scales (see volume 4 in this series). Best practice involves carrying a subset of items over from test to test to provide a strong means to link tests. It is also essential that student samples and the procedures followed in administration be equivalent. If exclusion criteria (for example, for students with learning difficulties) vary from one assessment to another, or if conditions over which administrators do not have control (for example, response rates) differ, such factors should be taken into account when comparisons are made between students' achievements at different points in time.

Type Of Assessment

The potential for use of the information derived from an assessment depends on the characteristics of the assessment. The use that can be made of assessment results varies for (a) census-based assessments, in which all (or most) schools and students in the target population participate (as, for example, in Brazil, Chile, and England); (b) sample-based assessments, in which a sample of students or schools that are selected to be representative of the total population take part (as is the practice in most countries); and (c) international assessments, in which a number of countries follow similar procedures to obtain information about student learning.

Census-Based Assessment

A national assessment in which all (or nearly all) schools and students, usually at specified grade or age levels, participate is termed census or population based. It has the potential to provide information on student achievement for (a) the education system in general, (b) sectors of the system, (c) schools, (d) teachers or classes, (e) individual students, and (f) factors associated with achievement. Because information is available about all schools, poorly performing schools can be identified, and

decisions can be made about possible interventions, such as the provision of teacher professional development courses, supplementary services, or additional resources. The assessment will become high stakes if sanctions are attached to school performance or if information about the performance of individual schools is published.

Sample-Based Assessment

Because the whole population does not participate in a sample-based assessment, it can, unlike a census-based assessment, provide information only on student achievement for (a) the education system in general, (b) sectors of the system, and (c) factors associated with achievement. Although this focus limits the use that can be made of the assessment's findings, it has a number of advantages. First, a sample-based assessment is considerably less expensive to administer than is a census-based one. Second, it is not necessary to assess all students to meet the basic objective of a national assessment, which is to provide valid, reliable, and timely information on the operation of the education system and, in particular, on the quality of student learning. Third, because participating schools are not identified, a sample-based assessment does not have the negative impact on schools and learning of a census-based assessment if sanctions for schools, teachers, or both are attached to performance. Finally, sample-based assessments can be administered more frequently, thereby allowing successive assessments to focus on emerging issues. Some national assessments are administered on an ongoing basis to rolling samples of students, thus giving educators access to assessment data on a continuous basis.

International Assessment

Another distinction that is relevant in considering the use of assessment data is whether the assessment is a stand-alone operation or is carried out in the context of an international study. International studies hold the promise of providing information that is not obtainable in a national assessment. They can (a) help define what is achievable (how much students can learn and at what age) by observing performance across a range of education systems; (b) allow researchers to observe and characterize the consequences of different practices and policies; (c) bring to light concepts for understanding education that may have been overlooked in a country; and (d) help identify and question beliefs and assumptions that may be taken for granted (Chabbot and Elliott 2003). Furthermore, international studies tend to achieve much higher technical standards than do national assessments, and they allow participants to share development and implementation costs that might otherwise put these methods out of reach in many systems. The results of international assessments tend to attract considerable media attention and have been used to fuel debate about the adequacy of educational provision and student achievement, as well as to propose changes in curricula (particularly in mathematics and science) (Robitaille, Beaton, and Plomp 2000). Although international assessments can—at least at a superficial level—provide comparative data on student achievement that are not available in a national

assessment, caution is necessary when it comes to using the findings to inform domestic policy. Among the potential pitfalls in using international data for this purpose is that because a test has to be administered in several countries, its content may not adequately represent the curriculum of any individual participating country. It is also generally recognized that international studies do not pay sufficient attention to the contexts within which education systems operate. Indeed, it is unlikely that the technology that they use can represent the subtleties of education systems or provide a fundamental understanding of learning and how it is influenced by local cultural and contextual factors (Porter and Gamoran 2002; Watson 1999). If so, then one cannot assume that approaches identified in international studies that appear to work well in some education systems will be equally effective in others. Not only might the adoption and implementation of policies based on this assumption be ineffective but they could actually be harmful (Robertson 2005). As well as providing comparisons between conditions in one's own education system and conditions in other systems, the data obtained in an international assessment may be used by an individual country to examine in-depth aspects of its own system (based on within-country analyses) in what becomes, in effect, a national assessment (Kuwait Ministry of Education 2008; Postlethwaite and Kellaghan 2008) (see box 1.6). Indeed, one of the aims of the International Association for the Evaluation of Educational Achievement 1990/91 Study of Reading Literacy was to provide national baseline data on the reading literacy of 9- and 14-year-olds for monitoring change over time (Elley 1992).

Underuse of National Assessment Findings

In considering the use of (and failure to use) national assessment findings, one must recognize at the outset that not a great deal of information is available about this topic. Furthermore, much less is available about the optimal use of findings or about the effects of basing policy decisions on the findings. This lack of findings may not be a true reflection of actual use, of course, because information related to use by government bodies may not be publicly documented. The evidence that is available indicates that the use of national assessment findings is not widespread, despite the potential that information derived from an assessment has for sparking reform and despite the expense incurred in obtaining such information. This observation has been made, for example, in the context of determining specific policies and decisions (see Arregui and McLauchlan 2005; Himmel 1996; Olivares 1996; Rojas and Esquivel 1998) and suggests that the use of national assessment data is very similar to the use of the findings of other policy-related research (see chapter 4). A description of the Honduran experience is probably typical of experience elsewhere (box 1.7). However, though identifying a specific use of assessment data might not have been possible in that case, the fact that the data influenced public opinion and raised consciousness is itself significant. A variety of reasons may be advanced for the underuse of national assessment findings (table 1.1). First, findings are likely to be underused when the national assessment is considered to be a stand-alone activity, separate from and with little connection to other educational

activity. This situation is likely to arise when national assessment activity is new or when it is carried out by external agents or at the request of donors. Rust (1999), for example, has pointed out that in Sub-Saharan Africa policy documents are often perceived by local bureaucrats as belonging to the donor agency and as separate from local policy making. Second, underuse of national assessment findings is likely to occur when policy makers, education managers, and other stakeholders who are in a position to act on findings have had limited or no involvement in the design and implementation of an assessment. Third, it is surprising, given the fact that assessments convey important information, that the first stage of use--the communication of information to relevant actors such as policy makers, providers of teacher training, and donors--is not always completed in a satisfactory manner, thus clearly limiting the potential for use. This problem may be due to a failure to budget for the dissemination of findings. In a situation in which most of the available project time and resources are required for the development and administration of instruments and analysis of data, nothing may have been left for the production and dissemination of information products and services. Fourth, the deficiencies of many assessments in instrumentation, sampling, and analysis can raise questions about the validity of the data they provide, causing potential users to pause before acting on findings or to dismiss the findings altogether. Fifth, if a national assessment identifies socioeconomic and educational inequalities associated with ethnic, racial, or religious group membership, this result may be a source of embarrassment to politicians, leading to attempts not to make findings public. Sixth, appropriate policy and managerial decisions are unlikely to ensue from a national assessment if procedures and mechanisms are not in place (a) to consider the findings in the context of other policy and managerial activities and (b) to determine action on the basis of assessment findings. Finally, national assessment findings are likely to remain underused unless all stakeholders who are in a position to act on findings (a) are adequately informed of the findings, (b) assess the implications of the assessment findings for their work, and (c) devise strategies designed to improve student learning. For example, in the case of schools and teachers, unless steps are taken to frame national assessment findings in a way that relates to teachers' concerns and unless funds are provided to create mechanisms by which teachers can use the information derived from an assessment to guide reform then the course of least resistance for school personnel may be at best to ignore the national assessment and at worst to undermine it. These observations should serve to caution against having unrealistic expectations for the policy changes that can follow an assessment. Nevertheless, this book tries to show that assessment data can provide guidance to policy and decision makers by elaborating on the actions designed to address underuse that are listed in table 1.1. When possible it cites examples, garnered from a large number of countries, of actual use both to arouse public interest and in the initiatives of policy makers and managers. Less evidence is available about the critical and more complex area of use of national assessment findings in classrooms to improve student learning.

BOX 1.6

Two Uses of Data from an International Assessment

The distinction between the use of data collected in an international assessment for cross-national comparisons and that of data collected for national analysis may be illustrated by a comparison of the use of data collected in International Association for the Evaluation of Educational Achievement (IEA) studies in Hungary and Finland. In Hungary, the focus was on a comparison between its system of education and systems in other countries. The comparison indicated that (a) achievements in mathematics and science were satisfactory, (b) achievements in reading comprehension were generally inferior to achievements in other countries, (c) schools in Hungary were more diverse in their achievements than schools in other countries, and (d) home background had a greater influence on students' reading achievement in Hungary than in other countries (Bóthory 1992). By contrast, in Finland, international data obtained in IEA studies were used to review the quality of mathematics and science education in the country. The data, which constituted the only empirical evidence available at the time on the achievements of students, were used in analyses that addressed particular needs and were fed to planners, policy makers, and ad hoc national working groups.

Source: Howie and Plomp 2005; Leimu 1992.

BOX 1.7

Extent of Use of National Assessment Results, Honduras

"These projects have produced a great amount of data and information describing students' educational achievement; they are research efforts which have a significant impact on public opinion but . . . have not contributed to make the educational system more efficient and effective. The results are scarcely used; they haven't been a real mechanism for control and accountability; and, to date, the information generated doesn't seem to have had meaningful consequences beyond informing, sensitizing, and raising consciousness."

Source: Moncada and others 2003: 73, as reported in Arregui and McLaughlan 2005: 6.

Table 1.1

Reasons for the Underuse of National Assessment Findings, Actions to Address Underuse, and Agents Responsible for Action

Reason	Action	Agent
1. National assessment activity regarded as a stand-alone activity, with little connection to other educational activities	Integrate assessment activity into existing structures, policy, and decision-making processes.	Ministry of education
2. Inadequate involvement of stakeholders in design and implementation of an assessment	Involve all relevant stakeholders in design and implementation of an assessment.	National assessment agency; ministry of education

3. Failure to communicate findings to all in a position to act	Make provision in the budget for dissemination, plan activities, and prepare a number of reports tailored to user needs.	National assessment agency
4. Lack of confidence in the findings of a national assessment	Ensure that the assessment team has the required technical competence and that relevant stakeholders are involved from the outset.	Ministry of education
5. Political sensitivity to making findings public	Increase the likelihood of making findings public by holding regular stakeholder discussions.	Ministry of education
6. Failure to devise appropriate action following an assessment at the level of general policies	Integrate national assessment activity into policy and managerial activities, and review findings to determine implications and strategies.	Ministry of education
7. Failure to devise appropriate action following a national assessment at the school level	schools, review findings and devise strategies to improve student achievement, and provide ongoing support for implementation.	National assessment agency; ministry of education; schools and teachers; teacher trainers; curriculum authorities; textbook providers

Conclusion

The use that can be made of the findings of a national assessment depends on a number of factors. The political context in which the assessment is carried out will have a strong bearing on use. Recognition that the assessment itself may be considered a political act reflecting the power, ideologies, and interests of social actors can serve to make the assessment and decisions based on it more transparent. Because the instrument used to measure students' achievement is the cornerstone of a national assessment, its quality will affect the use that can be made of findings. For optimum use, test instruments should provide information about student achievements (a) that is accurate and comprehensive, (b) that measures a range of achievements, (c) that provides guidance for remedial action, and (d) that is sensitive to instructional change. The tests used in many national assessments do not meet these conditions. They may be limited to measuring lower-order levels of knowledge and skills, they may not contain a sufficient number of items, and they may be too difficult, with the result that potential users do not have a reliable basis for policy and decisions. The value of a

national assessment for potential users will be enhanced if the background data on students' experience that are collected and the procedures that are used to analyze data--point to factors that affect student learning and are amenable to policy manipulation. A key decision for policy makers and education managers contemplating a national assessment that has implications for the use that can be made of findings is whether the assessment should be sample based or census based. A sample-based assessment will provide information and a basis for action at the system level, whereas a census-based one will, in addition, provide information about--and a basis for action in individual schools. The choice of a sample-based or census-based assessment should be guided by a consideration both of the information needs of policy makers and managers and of the cost involved. A census-based assessment provides the opportunity to hold schools accountable for student learning. Before deciding to use assessment findings for this purpose, policy makers should give serious consideration to (a) the limited information that a national assessment can provide about the quality of education provided by a school; (b) the range of individuals, institutions, and conditions that affect student learning; and (c) the negative (if unintended) consequences of attaching high stakes to student performance. Although an assessment used in this way as a mechanism of power may be corrective in the short term, in the longer term the bureaucratic imperative associated with it may corrupt the system that it was designed to correct or improve (Madaus and Kellaghan 1992). When significant direct consequences are not attached to results, which is the case in most national assessments, the assessment is considered low stakes, and findings will be used primarily as a tool for planning and management (McDonnell 2005). The information that is obtained is considered to be a sufficient incentive for politicians, policy makers, educators, parents, and the public to act, and though the state may not accept responsibility for actual student achievement, it does accept its responsibility to make adequate provision for public education and to reduce disparities in the quality of education offered to--and achieved by--children of different ethnic backgrounds or social classes (Reimers 2003). When a state adopts this position, detailed analysis of test results will be required to describe student achievements and to identify school and teacher practices that enhance those achievements. Following this, findings should be widely disseminated, resources and technical assistance should be provided to help schools identify problems they are experiencing, and continuing support should be provided for a process of school improvement. This series of books has been written primarily to serve the needs of individuals carrying out a sample-based national assessment. However, the content of other volumes, except the module on sampling and some of the statistical analysis module, is also relevant to implementation of a census-based assessment. Much of the present volume is also relevant, though a number of issues are not (for example, identification of schools in need of assistance following a sample-based assessment). The prerequisites for effective use of the findings of a national assessment that will be considered are relevant to both sample- and census-based assessments and include the following:

- Involving policy and decision makers in the design of the assessment to address issues that they have identified as of pressing interest
- Communicating results in a timely fashion and in a form that is intelligible to key users
 - Incorporating assessment information into existing bureaucratic structures and translating such information into policy, strategies, and policy instruments (for example, mandates, capacity-building strategies, inducements, and hortatory policies to motivate action)
 - Ensuring that assessment findings influence the practice of classroom teachers, with the objective of improving student learning
 - Providing continuing political support to use the findings to bring about change and to devise mechanisms that support their application in reform at the classroom level. Throughout the volume, as the many activities that a national assessment can spawn are described, reference is made to census-based and international studies when they provide insights to use or when they describe practices that are relevant to a sample-based assessment. Chapters 2 and 3 describe the types of reports that are needed to inform users of the findings of an assessment. Chapter 4 outlines general issues that merit consideration when translating assessment findings into policy and action. This chapter is followed by a description of specific uses of national assessment data for policy and educational management (chapter 5), for teaching (chapter 6), and to promote public awareness (chapter 7). The concluding chapter (chapter 8) identifies conditions that are likely to optimize use of the findings of a national assessment. It also suggests a number of ways in which national assessment activities could be modified and enhanced with a view to increasing their value to users.

Chapter 2

Reporting a National Assessment: the Main Report

This chapter outlines the components of the main and essential instrument for reporting the findings of a national assessment. These components should include not only the findings but also the procedures followed throughout the assessment so that readers can judge their adequacy and relevance. The report will also form the basis of ancillary means of communicating the findings (for example, briefing notes, press releases, a report for schools – see chapter 3). The main report of a national assessment should contain a description of the following components: (a) context of the assessment, (b) objectives of the assessment, (c) framework that guided the design of the assessment, (d) procedures followed, (e) descriptions of achievement in the national assessment, (f) correlates of achievement, and (g) changes in achievement over time (if appropriate data are available from a number of assessments). The amount of detail presented in the main report depends on whether a separate technical report is prepared. Most readers will have limited technical

knowledge and are interested only in what the report implies for their work. Much of the technical detail can be assigned to the main report's appendixes. At the outset, members of the national assessment team and key stakeholders should generally agree on how to design the main report, collect the data, and report the results. Reaching agreement about reporting results can be facilitated by drafting a series of blank or dummy tables and discussing the precise variables and data associated with each table. Table 2.1 is a blank table used to illustrate how national student level data might be presented by curriculum area and gender. Table 2.2 suggests how provincial-level results might be presented to allow policy makers to compare levels of achievement among low-achieving students (those at the 5th percentile) and high-achieving students (those at the 95th percentile) in each province. Table 2.3 compares students' level of achievement at two points in time. Table 2.4 is designed to identify relationships between student achievement and a number of variables of interest to policy.

Table 2.1

Mean Scores (and Standard Errors) of Boys and Girls in a National Assessment of Language and Mathematics

	Language		Mathematics	
	Boys	Girls	Boys	Girls
Mean				
Standard error				

Source: Authors' representation.

Table 2.2

Mean Scores (and Standard Errors) and Scores at Varying Percentile Ranks in a National Assessment of Science, by Province

Province	Mean	Standard error	Percentile score			
			5th	25th	75th	95th
1						
2						
3						
4						
5						
6						
National						

Source: Authors' representation.

Note: These data can be used to prepare a box and whisker-type plots.

Table 2.3

Mean Achievement Scores (and Standard Errors) in a National Assessment Administered at Two Points in Time

First administration		Second administration		Statistically significant?
Mean	Standard error	Mean	Standard error	

Source: Authors' representation.

Note: One must take into account that both means are sample based in calculating the significance of the difference between them.

Table 2.4

Correlation between Mean School Reading Achievement Scores and School Factors in a Grade 5 National Assessment

Variables	r
Average class size, grade 5	
Average number of textbooks per student, grade 5	
Percentage of students who can be seated during writing class, grade 5	
Number of classroom teachers in school	
Number of registered students in school	
Number of years of teaching experience (grade 5 teachers)	
Level of teacher qualification (grade 5 teachers)	
Number of classroom resources	

Source: Authors' representation.

Context of the National Assessment

In describing context, one may state the importance of obtaining information on student learning as a basis for policy and management decisions. A consideration of evidence from earlier studies on students' achievements (if available) will be relevant.

Objectives of the National Assessment

The main objective should be stated: for example, to provide evidence on student learning in the education system. More specific objectives may also be stated: for example, to establish the current reading standards of fourth-grade pupils; to compare student achievements in private and public schools; to monitor trends in student learning over time; to describe school resources; to examine school, home background, and pupil factors that may be related to reading achievement; and to provide a basis for future assessments.

Framework for the National Assessment

A framework is an overall plan or outline that describes what is being assessed in terms of knowledge, skills, and other attributes and how it is being assessed. The framework guides the development of the assessment and makes the assessment transparent, first, for those who construct the assessment instruments, but also for the wider audience who will read the report of the assessment. Chapter 2 in volume 2 in this series describes how to develop an assessment framework (Anderson and Morgan 2008). The Progress in International Reading Literacy Study (PIRLS) provides an example of a description of the construct assessed in its study of the reading achievements of nine-year-olds (Mullis and others 2006; see also volume 1 of this series, Greaney and Kellaghan 2008: appendix B2). Reading is described in terms of two purposes (reading for literary experience and reading to acquire and use information) and four processes (focusing on and retrieving explicitly stated information, making straightforward inferences, interpreting and integrating ideas and information, and examining and evaluating content). A framework also describes the instruments used to assess achievement. Including examples of the types of item used in the assessment is useful to provide readers with an idea of the nature of the tasks involved. Of course, these items should not include those planned for use in future assessments.

Procedures in Administration of the National Assessment

How and when data were collected should be described. This description will include identification of the population on which the assessment was based, selection of schools or students for participation, and data on exclusions and nonparticipation.

Description of Achievement in the National Assessment

In deciding how to present the findings of a national assessment, it is important to bear in mind that the information provided should be relevant to policy makers' and decision makers' needs and should assist them in addressing policy problems constructively. The choice of a single index of student achievement (for example, a total mathematics score) or multiple indexes (for example, separate scores for computation and problem solving) may be relevant. Although policy makers may generally prefer summary statistics, reporting only a single index of achievement will most likely miss important information, thereby limiting the basis for action following the assessment (Kupermintz and others 1995). Increasingly, a description of performance in terms of proficiency levels is being used to present the results of a national assessment. The procedure involves scale anchoring, which has two components: (a) a statistical component that identifies items that discriminate between successive points on the proficiency scale using specific item characteristics (for example, the proportions of successful responses to items

at different score levels) and (b) a consensus component in which identified items are used by curriculum specialists to provide an interpretation of what groups of students at, or close to, the related points know and can do (Beaton and Allen 1992). The levels may be labeled (for example, satisfactory/unsatisfactory; minimum/desired; basic/proficient/advanced), and the proportion of students achieving at each level identified. Table 2.5 presents data from a national assessment in Mauritius. Table 2.6, which describes levels of mathematics achievement in the U.S. National Assessment of Educational Progress (NAEP), goes beyond the type of data in table 2.5 in providing definitions of performance at a range of proficiency levels. The percentage of students (in public schools) at each level ranged from 44 percent at the basic level, to 30 percent at the proficient level, to 5 percent at the advanced level. Thus, 79 percent of students performed at or above the basic level (Perie, Grigg, and Dion 2005).

Table 2.5

Percentages of Students Scoring at Minimum and Desired Levels of Mastery in Literacy, Numeracy, and Life Skills Tests: Mauritius

Subject	Percentage of students	
	At or above minimal level	At or above desired level
Literacy	77.6	35.4
Numeracy	70.3	26.4
Life skills	71.6	32.4

Source: Mauritius Examinations Syndicate 2003. Reproduced with permission.

Table 2.6

NAEP Mathematics Achievement Levels, Grade 4: United States

Level	Expected achievement at grade 4
Basic	Students should show some evidence of understanding mathematics concepts and procedures in the five NAEP content areas. They should be able to estimate and use basic facts to perform simple computations with whole numbers, show some understanding of fractions and decimals, and solve some simple real-world problems in all NAEP content areas. They should be able to use, although not always accurately, four-function calculators, rulers, and geometric shapes. Their written responses will often be minimal and presented without supporting information.

Proficient	Students should consistently apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas. They should be able to use whole numbers to estimate, compute, and determine whether results are reasonable. They should have a conceptual understanding of fractions and decimals; be able to solve real-world problems in all NAEP content areas; and use four-function calculators, rulers, and geometric shapes appropriately. They should employ problem-solving strategies such as identifying and using appropriate information. Their written solutions should be organized and presented with both supporting information and explanations of how the solutions were achieved.
Advanced	Students should apply integrated procedural knowledge and conceptual understanding to complex and nonroutine real-world problem solving in the five NAEP content areas. They should be able to solve complex and nonroutine real-world problems in all NAEP content areas. They should display mastery in the use of four-function calculators, rulers, and geometric shapes. They are expected to draw logical conclusions and justify answers and solution processes by explaining why, as well as how, the solutions were achieved. They should go beyond the obvious in their interpretations and be able to communicate their thoughts clearly and concisely.

Source: U.S. National Center for Education Statistics 2006a.

The approach to establishing proficiency levels differs in the 2001 PIRLS. Cut points were determined first by specifying the percentage of students in each benchmark category and then by examining the reading skills and strategies associated with each level (figure 2.1). Vietnamese policy makers, working with other interested parties such as curriculum developers, identified six levels of student achievement in reading for grade 5 students using statistical information and the judgments of experts (table 2.7). Policy makers used the data to make national, provincial, and other comparisons of achievement. In many national assessments, variance in student achievement is partitioned into between- and within-school components. This process involves calculating the intraclass correlation coefficient (ρ), which is a measure of the homogeneity of student achievement within schools. It tells how much of the variation in achievement is between students within schools (within clusters) and how much is between schools (between clusters). A low intraclass coefficient means that schools perform at comparable levels, while increasing values of the coefficient indicate increasing variation between schools in student achievement (Postlethwaite 1995). The findings of international studies (for example, PIRLS or the Programme for International Student Assessment, known as PISA) indicate that considerable differences exist between education systems in the value of the intraclass correlation. Furthermore, systems in which the national level of achievement is low tend to exhibit greater differences between schools in their achievement levels.

Table 2.7
Grade 5 Reading Skill Levels in National Assessment: Vietnam

Skill level	Achievement	Percent of students	Standard error
1	Student matches text at word or sentence level, aided by pictures. Skill is restricted to a limited range of vocabulary linked to pictures.	4.6	0.17
2	Student locates text expressed in short, repetitive sentences and can deal with text unaided by pictures. Type of text is limited to short sentences and phrases with repetitive patterns.	14.4	0.28
3	Student reads and understands longer passages. Student can search backward or forward through text for information and understands paraphrasing. An expanding vocabulary enables understanding of sentences with some complex structure.	23.1	0.34
4	Student links information from different parts of the text. Student selects and connects text to derive and infer different possible meanings.	20.2	0.27
5	Student links inferences and identifies an author's intention from information stated in different ways, in different text types, and in documents where the message is not explicit.	24.5	0.39
6	Student combines text with outside knowledge to infer various meanings, including hidden meanings. Student identifies an author's purposes, attitudes, values, beliefs, motives, unstated assumptions, and arguments.	13.1	0.41

Текстът продължава в кн.2/2014 г. на сп. „Стратегии на образователната и научната политика“

The text continues in issue 2/2014 of “Strategies for Policy in Science and Education Journal”