Large-Scale Student Assessment: Guidelines for Policymakers

Charles Ungerleider

Department of Education Studies
University of British Columbia

Targeted for policy-makers, this article provides suggestions for ensuring that the benefits of using large-scale student assessments are achieved in the face of a number of challenges to their effective use. Fundamental concepts and concerns are reviewed in the context of making valid inferences from test scores. A unique feature of this article is that the author draws on his recent experiences in the educational policy arena.

Despite their utility for (a) determining whether students are meeting the standards set for their achievement and (b) for providing information for improving student achievement, large-scale assessments are not typically used well. The purpose of this article is to provide policy-makers with suggestions for ensuring that the benefits of using large-scale student assessments are achieved in the face of a number of challenges to their effective use (cf. Taylor & Tubianosa, 2001; Kohn, 2000).

This article addresses major challenges to the effective use of large-scale student assessments, but does not address many issues affecting the construction of the assessments to ensure their psychometric integrity.

One challenge is to distinguish between norm- and criterion-referenced assessment. Norm-referenced assessments are primarily aimed at determining how well students perform in relation to one another. Criterion-referenced assessments are primarily aimed at determining which students have achieved mastery of a body of knowledge deemed appropriate for the students who take the test, although they are occasionally used to make comparisons among students. This article addresses the use of criterion-referenced large-scale student assessments (CRLSAs) as
means of (a) determining whether students are meeting the standards set for their achievement and (b) providing information for improving student achievement.

Another challenge is ensuring a match between the assessment and the curriculum that students are expected to have mastered by the time the assessment is conducted. The effective use of the results from CRLSAs depends on clearly articulated curricula, specific learning outcomes, and widely accepted curricular exemplars. The curricula should specify clearly the knowledge that students are expected to master and when they are expected to have it mastered. It is essential to fully explore and gather curriculum exemplars of acceptable student work, student work that exceeds expectations, and work that falls short of the desired standard for the construction of assessment instruments. These pieces of work are also helpful to teachers and parents in understanding the relation between student performances and the standard. It is desirable that jurisdictions ensure the participation of classroom teachers in the curriculum development process and, in particular, the identification of relevant curriculum exemplars. Their experience with and knowledge of the student population for whom the curriculum is intended will help to ensure that the criteria developed are appropriate.

The involvement of classroom teachers is also an important part of successful assessment programs. In addition to the importance of such involvement for ensuring a match between the assessment and curriculum, the participation of teachers from the relevant jurisdiction at all stages of the assessment process (i.e., development, field-testing, administration, marking, standard-setting, interpretation) is critical to gaining trust and grass-roots support. Though many Canadian jurisdictions involve classroom teachers, many departments of education in the United States hire companies to develop, mark, and interpret the results of their tests. Involvement of local teachers in all aspects of the assessment process is beneficial for everyone. Teachers benefit professionally by learning about assessment practices and from their collaboration with other teachers from elsewhere in the jurisdiction. Ministries of education benefit from the development of valid and better-respected assessment instruments.

Teachers faced with centralized testing regimes feel that they must ensure that students have mastered the knowledge that will be assessed. If the time available for instruction is not sufficient to achieve the intended learning outcomes for all students, teachers confronted with CRLSAs may sacrifice other important aspects of the domain being assessed or “borrow” time from curricular areas that are not assessed. Thus, it is important that the required learning outcomes that will be formally assessed using CRLSAs are clearly indicated, are achievable in the time allocated for instruction, and can be achieved without sacrificing other important learning outcomes.

CRLSAs depend on teachers having appropriate instructional material. In addition to having sufficient quantities of instructional material, it is important that the instructional material available be closely linked to the curriculum as it is intended
to be implemented and learned. If knowledge is of sufficient importance to ensure its mastery through CRLSAs, it is essential that the knowledge be addressed well in the instructional support material.

Credibility of the results can be an obstacle if the standards themselves are not widely accepted. Determination of standards of mastery must be made prior to assessment construction and should involve parents, teachers, administrators, and representatives of the broader public. Policy-makers must be prepared for the increased time that will be required for standard setting and preparation of the assessment as well as for the conflict that will likely ensue when expectations of different constituencies differ. It is likely that business people, university educators, teachers, and parents will hold different perspectives about standards. Though possible and desirable, obtaining agreement about standards with the involvement of a wider constituency will be more difficult than if only grade-level teachers or subject specialists are involved. Jurisdictions that do not involve a range of educational partners in the standard setting process risk that the standards will not be respected by all partners.

When standards are used in reporting the results of large-scale assessments, it is important to ensure that the standards remain relatively consistent from one year to the next across successive administrations. Jurisdictions that want to assess growth over time using different tests will need to employ proper test equating procedures (e.g., through use of a common set of anchor items and item response theory) so that the results of different tests in different years can be placed on a common metric and compared to a common standard. Use of these types of test equating procedures will also help address problems relating to the creation of inconsistent standards across different standard-setting panels. Test equating also eliminates the need for annual standard-setting activities because the same standards could be applied across a number of years of assessments, with periodic reviews every 4 or 5 years to ensure that the standards reflect current expectations. If the assessment addresses broadly accepted and important dimensions of learning, the standards should not change dramatically over time.

Policy-makers should be aware that some of the planned consequences resulting from centralized testing regimes might have a deleterious impact on teaching and learning. The decline of promotion of students from one grade to the next, teacher and administrator reassignment, and school closures are examples of “high stakes” consequences of assessment (cf. Firestone & Mayrowetz, 2000). The prospect of such consequences have produced a variety of undesirable behaviors, including coaching students while the assessment is being conducted or inappropriately excluding students whose performances are likely to negatively influence the results. Even when the stakes of assessments are seemingly nonexistent, there are unintended consequences that must be taken into account (cf. McNeil, 2000). Knowing that the results do not “count,” students may not put much effort into their responses. Policy-makers should avoid either extreme. Making the assess-
ment part of the final grade students receive in the subject assessed should ensure that students take the assessment seriously.

The work of Coleman and his associates (1966) helped to sensitize educators to the close association between the socioeconomic background of youngsters and their achievements. The use of CLRSA has been challenged on the grounds that the results from centralized testing regimes can normally be predicted from the background characteristics of the students whose performances are being assessed.

State and district monitoring systems usually report school mean test scores, or the percentage of students achieving some criteria, without regard to the characteristics of students entering the school, measurement and sampling error, or the wider social and economic factors that contribute to schooling outcomes. This practice has lead to widespread misinterpretation of data on school achievement, because higher test scores tend to be mistakenly equated with more effective schooling. It has also created mistrust and scepticism among many educators, such that there is widespread resistance to any kind of performance monitoring. (Willms, 1998, p. 4; cf. Kohn, 2001)

The utility of CLRASAs for improving student achievement depends on the capacity for investigating relations among variables over which the system exercises control or is capable of exercising control. CLRASAs depend on the ability to use advanced statistical techniques such as hierarchical linear modelling to investigate relations (Hao, 1999; Lee, 2001). Provincial and state departments of education should be able to ensure the integration of the data from CLRASAs with other relevant data sets and to provide the analysis and interpretation necessary to isolate the factors over which schools have control. The ability to identify classroom and school attributes and relate them to results will be crucial to making full use of CLRASAs. In a sense, CLRASAs also depend on building research capacity and interest. Making data available to qualified researchers will encourage interest and stimulate capacity for analyzing large sets of data with appropriate statistical techniques.

Those contemplating the use of CLRASAs must make provision for disentangling the influence of nonschool variables on the achievements of students: that is, discerning a school’s effects from other effects (e.g., socioeconomic background, ethnicity, and gender). Willms (1998) argued persuasively for the use of socioeconomic gradients, depicting the relation between the socioeconomic background of the students and the outcomes being assessed, and for multivariate analyses using sophisticated statistical techniques. Citing the advantages that accrue as a consequence of having schools or classrooms with students from advantaged backgrounds or who possess high initial ability, Willms (1998) argued that it is also necessary to take into account the school’s contextual effects, distinguishing between such factors as the socioeconomic profile of the school community and the effects specifically attributable to teaching (e.g.,
the use of strategies that maximize the achievement of all students) and school policies and practices such as educational leadership, academic expectations, curricular focus, disciplinary climate, and parental involvement (see, e.g., Ma & Klinger, 2000). From an educational accountability standpoint, it is the effects that are attributable to teaching and school policies and that are of greatest interest. As difficult as it may be to ascertain the effects of teaching and school policies and practices, it is necessary to isolate them from effects of student background (e.g., socioeconomic status, gender, ethnic group membership, etc.) and aptitude (e.g., prior ability, attitude toward school, locus-of-control, etc.), as well as the contextual effects of attending a particular school (Raudenbush & Willms, 1995).

Expressing the relationship between school attributes over which policy-makers exercise control and student outcomes in familiar terms is essential in presenting the data to provincial and local politicians. For example, it should be possible to make such statements as “one can reduce the proportion of students scoring below grade level by reducing class sizes at the primary level from X to Y” (cf. Cohen, Raudenbush, & Ball, 2000).

The maximum benefits of CRLSAs accrue by tracking the performance of specific students over time and computing the school’s impact on their growth trajectory (Willms & Audas, 2000). Cohort studies are of more limited use in determining student improvement than studies that track and measure the performance of students at various stages in their school careers. Growth trajectories can be computed from assessments conducted; for example, at Grades 4, 7, and 10 (cf. Bloom, 1999).

As Runte (1998) made clear, the potential for misusing the results of CRLSA is great. Though intended primarily for professional use and for informing parents about the progress of their children, others—including education professionals, local and provincial politicians, and the media—may unintentionally or wilfully misuse the results of the assessments. Teachers are not likely to embrace the use of CRLSA as an aid to instructional decision making if the environment is punishment centered. Comparisons among classrooms, schools, or districts without proper regard for and control of situational or other differences between them are likely to exacerbate teacher opposition to the use of such information. Administrators, school trustees, parents, and politicians who misuse the results as a measure of teacher performance or productivity are not likely to engender teacher confidence in the process. Invidious comparisons of classrooms, schools, and districts abound in popular media. Politicians and educators who should know better make inappropriate comparisons. Jurisdictions contemplating the use of CRLSAs must plan carefully for the dissemination of results, prepare professionals, journalists, and politicians for their appropriate use, and deliberately counter misuse of the data.

Effective use of the information provided by CRLSAs depends on a school environment that fosters a sense of collective responsibility for the educational suc-
cess of students. Much of the variation in student achievement that is amenable to intervention occurs at the classroom level. Instruction is the principal, but not sole, mediating link between assessment and student achievement; other variables include the disciplinary climate of the school and the school’s emphasis on academic achievement. Effort and resources should be focused on reducing the variation among students in general and between identifiable groups of students in particular. CRLSAs cannot succeed in the face of organizational factors that impede the accomplishment of the required learning outcomes. These factors include inadequate time allocation (i.e., hours of instruction), inappropriate time distribution. (i.e., timetabling of subjects), and insufficient time for teachers to review collaboratively the assessment data and to consider what, if any, modifications should be made to their instructional plans, to the methodology they employ in the classroom, or to school policies and practices.

Inadequate preparation of teachers is an obstacle to the effective use of CRLSAs. Faculties of education should ensure that those preparing to teach possess the knowledge they will need in the use of assessment information for instructional planning and implementation. Generalized measurement courses do not typically develop the knowledge one needs to use information from CRLSAs for instructional planning and implementation. Such knowledge should be developed in the context of curriculum and instruction courses and should complement the knowledge about teacher-designed assessment included in those courses.

Continuing professional education should be provided for experienced teachers in the use of assessment information for instructional planning and implementation. Responsibility for providing such continuing education falls to the jurisdiction that is responsible for the assessment. In Canada and the United States, provinces and states typically promulgate CRLSAs. These jurisdictions must also provide for the professional education that practitioners require.

The lack of school-based leadership can be an obstacle to the use of the information provided by CRLSAs. The importance of school-based leadership has been well established, though less well developed in practice. Insufficient attention has been given to instructional leadership ability in the recruitment, selection, and preparation of school-based administrators. As a consequence, school-based and district administrators are often ill-disposed and ill-equipped to lead the staff in carefully considering the results of CRLSAs and the implications of the results for school policies and practices and for instruction. If improving student success is to remain a central goal of education, administrators must be steeped in the intricacies of using information from CRLSA. Possessing such capacity must become a prerequisite for holding an administrative position. Responsibility for providing continuing professional education for administrators currently employed falls to the jurisdiction responsible for promulgating such assessments.
The absence of close working relationships among the agencies and organizations responsible for the various dimensions of education and educational improvement will hamper the appropriate and most efficacious use of CRLSAs. The agencies responsible for certification of teachers and administrators must work closely with faculties of education and the other agencies providing for professional education. Provincial and state departments and ministries of education need to work with school boards, senior administrators, teacher organizations, and parent groups to facilitate understanding of and support for the appropriate use of CLRSA. School boards must monitor the information provided by CRLSAs and take it into account in their performance plans. Ensuring school success for all students requires jurisdictions to distribute resources to ensure equality of outcomes, especially for those students for whom learning is more challenging because of poverty and discrimination.

CRLSAs do not by themselves improve the performance of students. Without making that a large number of other factors are present, it would be unrealistic and unsound to embark on such assessment. The use of large-scale assessments to inform changes at the school and classroom level must address the aforementioned obstacles and challenges. At a minimum, the following actions (after Willms, 1998; cf. Woessmann, 2001) appear to be essential:

- Establish broad agreement about what school outcomes are essential for all students.
- Ensure that these outcomes are clearly articulated in the curriculum and are supported with appropriate instructional material.
- Hold students, parents, and teachers accountable for those outcomes.
- Assess student progress in the areas of importance at different times over their school careers.
- Prepare teachers and encourage them to use teaching strategies that increase learning outcomes for all students.
- Encourage mixed-ability grouping and discourage grouping, tracking, or streaming students by socioeconomic background or in ways that increase differentiation among students of different ethnocultural backgrounds.
- Assess schools on the basis of student growth in learning outcomes, taking into account their individual socioeconomic backgrounds, the socioeconomic context of the school community, as well as school policies and practices known to influence the achievement of the valued outcomes.
- Examine rates of student progress as well as gradients in student progress associated with such background factors as socioeconomic standing, gender, and ethnicity.
- Ensure that teachers and administrators are well prepared for their responsibilities.
- Counter misuse of the results of CLRSAs in the media and elsewhere.
• Provide teachers with adequate time to individually and collectively interpret data for the purpose of improving instruction.

As I hope I have made clear, improving student success will depend as much on other factors as it does on the assessment system itself. Improving the capacity of teachers to employ teaching strategies that increase the performance of all students, developing the capacity of principals to provide instructional leadership, helping district administrators to acquire the capacity to interpret and use the information provided by large-scale student assessments will require thoughtful and focused professional preparation and continuing education. Educating school district or provincial politicians about the impact of their comments on the results of such assessments, and countering erroneous or wilfully misleading representations of the results of such assessments by the media or others, will be challenging and will require significant attention.

Despite the challenges and obstacles to their effective use, if the data from CRLSAs are subjected to the rigorous statistical analyses and interpretation, they can be very useful in improving the conditions influencing student achievement. Unless ministries of education are willing to analyze and interpret the data with care and rigor, I fear that large-scale student assessments will be used to bludgeon schools and teachers rather than to inform them about the changes that might be made to improve student success.

In the face of anxiety about the future of the next generation, and thus our own future, and in the absence of other publicly available standards, too many people—including ones who should know better—invest too much importance in the results of assessments and too little in the analyses and interpretation of those results. Policy-makers would do well to consider the differences between blunt and ideologically motivated analyses of data such as the Fraser Institute’s annual reports on secondary schooling (Cowley & Easton, 2001; Cowley & Shahabi-Azad, 2001a, 2000b) and the finer grained and more useful analyses of Ma and Klinger (2000) or Willms (1998), for example.

The analyses and interpretations of the data from large-scale measures of student achievement are best used on small-scale initiatives to improve student achievement. Consider how countries such as Canada and the United States use periodic reports on the fluctuations in their economies. They invest little import in short-term fluctuations, reacting only to patterns that develop over time. They seldom respond to the data by making system-wide economic changes. Instead, they direct interventions at specific areas of the economy. The problem with the large-scale assessments programs I have seen is that the data reported prompt calls for radical reform of the system rather than thoughtful consideration of classroom and school practices. Policy-makers need to get beyond viewing large-scale student assessments as “accountability” measures and embrace them as one, and only one, source of information that can be used to improve student success in school.
ACKNOWLEDGMENTS

This article is a response to an invitation I received when I was Deputy Minister of Education for the Province of British Columbia. I was asked to address the issue of large-scale student assessment from a “policy perspective.” In responding to that invitation, I bring my background as a sociologist of education, my observations of the use of large-scale assessments in other jurisdictions, and my relatively recent experience as Deputy Minister.

I am grateful to Dean Goodman, Jerry Mussio, David Robitaille, and Thomas Stephens for helpful suggestions for the improvement of the article. They are not, of course, responsible for any failure to follow their advice!

REFERENCES


