Harcourt

policy report

Toward Complete Assessment Systems: *Stanford Learning First*™

Margaret A. Jorgensen, Ph.D.

Sasha Zucker

November 2004 *(Revision 1, December 2004)*

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

Toward Complete Assessment Systems: Stanford Learning First

Introduction

In the United States, there is a wide range of educational assessment systems used for many different purposes. Policymakers, administrators, and other decision-makers rely on educational assessments to evaluate instructional programs, allocate resources, and guide reform. Parents rely on assessment results to gauge what their children are learning. Teachers use educational assessments in their day-to-day classroom activities to monitor student progress. Because of these different purposes, no assessment is "one-size-fits-all." Any assessment designed to serve a particular role in an education system has shortcomings when used for a purpose beyond its intended use. However, in a complete, ideal assessment system, all of the assessments are coherently designed to work together in support of student achievement. This report examines and discusses how a complete assessment program might incorporate a range of assessments to effectively address the needs of the participants in an education system.

Fundamental Classifications of Assessments

A review of the way in which education assessments are classified is illuminating to a discussion of the characteristics of a complete assessment system. Fundamentally, assessments can be classified according to 1) their purpose (*summative assessments* versus *formative assessments*) and 2) their scope of content (*classroom assessments* versus *accountability assessments*).

Purpose

Educational assessments can be used in summative or formative roles. A *summative assessment* is administered after a teacher has completed instruction in a subject area to determine what a student knows and can do. Common examples include end-of-course exams and accountability assessments that are administered to indicate a student's level of proficiency. In contrast, a *formative assessment* is administered periodically during instruction to monitor student progress. Teachers

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

use the results from formative assessments to meet the educational needs of their students by planning and changing instruction. An end-of-chapter test is one common example of a formative assessment. Explained succinctly, a summative assessment is an assessment *of* learning *after* instruction has taken place, while a formative assessment is *for* learning *during* instruction (Arter, 2004; National Research Council, 2003). While an assessment is typically designed to be used only for one purpose or the other, it is possible for an assessment to produce results that can be used for both roles.

Scope

Depending on the content coverage, educational assessments can be classified as either *classroom assessments* or *accountability assessments* (National Research Council, 2003; Pellegrino, Chudowsky, and Glaser, 2001). Classroom assessments include the broad range of tests developed and administered by teachers as a part of day-to-day classroom activities. Teachers use classroom assessments in both formative and summative contexts, such as planning instruction, monitoring the progress of students on a weekly or daily basis, and establishing a student's final grade in a course (Airasian, 2001; National Research Council, 2003; Nitko, 2004). Typically, classroom assessments have little or no empirical research supporting their design and use.

The *No Child Left Behind Act* of 2001 (NCLB), has increased the public awareness of accountability assessments. In contrast to classroom assessments, accountability assessments are developed using highly formal, standardized procedures that are guided by well-documented best practices and supported by decades of psychometric theory and scientific research (Nitko, 2004). Accountability assessments are administered to large groups of students, typically the student population of an entire state. Under NCLB, accountability assessments are used to produce scores that indicate a student's level of achievement compared to performance standards set for each subject area and grade level (National Research Council, 2003). While some states continue to invest in customized assessments (tests created by a particular state), others have chosen to buy and augment pre-existing published tests to maximize coverage of content standards. To the extent that the base product is a norm-referenced test, augmented tests rank a student's achievement compared to a national sample of students as well as indicating the student's achievement compared to standards. Augmented tests represent an efficient, economical solution for states to meet the testing mandate of NCLB while also obtaining useful comparative data (Jorgensen and McBee, 2003; National Research Council, 2003).

Under NCLB, accountability assessments are used in a summative context to determine whether a student has reached a required level of proficiency and whether a school's entire student population and disaggregated subgroups have

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

made Adequate Yearly Progress (AYP). If accountability tests are designed to report achievement on individual content standards, they also can be used in a formative context to determine relative strengths and weaknesses. The results obtained from an accountability assessment are frequently used by administrators and policymakers to evaluate a curriculum or instructional program and identify areas for improvement (National Research Council, 2003).

Examining Differences and Tradeoffs in Assessment Design

The purpose of accountability assessments and classroom assessments affects their respective designs. Both types of assessments have marked advantages when used in the role for which they are designed. However, gaps in their utility emerge when they are used in roles for which they are not designed. Examining the differences and tradeoffs made in the design of classroom and accountability assessments provides insights concerning how different types of assessments function together in a complete assessment system.

Qualities of Accountability Assessments

A clear strength of accountability assessments is the tremendous resources that are devoted to their development and administration. The far-reaching decisions that depend on the results of accountability assessments bring a high degree of publicity and scrutiny from the public. Hence, test publishers develop accountability assessments in adherence to widely accepted standards for reliability, validity, and fairness found in publications such as the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999). Publishers execute rigorous studies to establish norm groups, certify content as robust and free of bias, and provide evidence that accountability assessments are statistically valid, reliable, and sound. Because of this rigor, the results obtained from published assessments are considered objective and provide clear feedback about the achievement of students (Nitko, 2004).

To provide useful results, accountability assessments must also be *aligned* with the content standards of a state. In educational assessment, alignment is the process by which an assessment is made to agree with and cover the content of a state's academic standards in sufficient breadth, depth, and balance (Ananda, 2003). Webb (1997) refers to this type of alignment as "horizontal alignment" in contrast to "vertical alignment" (the differences between these two conceptions of alignment are discussed further in this report). Alignment is an important quality of any accountability assessment for both its intrinsic benefits, such as the test's validity, and its extrinsic effects, such as the test's ability to concretely exemplify a state's educational goals and standards to students and teachers (National Research Council, 2003; Nitko, 2004).

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

The design of accountability assessments also results in some tradeoffs. While accountability assessments produce objective measurements of levels of achievement across a subject, such as proficiency in reading, they are usually not precise enough in their focus to inform instruction. Teachers do not always have sufficient training to use the results of accountability assessments in a formative context. Moreover, the reporting timeline for accountability assessment results, typically several weeks, may preclude their immediate usefulness in guiding and improving instruction. However, these tradeoffs may be acceptable to administrators who have informational needs that are quite different from those of teachers (National Research Council, 2003).

Qualities of Classroom Assessments

Because classroom assessments are part of daily instruction, the principal advantage of classroom assessments is the immediate feedback they provide to teachers and students (Nitko, 2004). Recent scientific research in the United States and Great Britain indicates that student achievement can be improved through the use of classroom assessments that provide timely corrective feedback (Marzano, Pickering, Pollock, 2001; National Research Council, 2003).

However, some characteristics of classroom assessments may lead to erroneous conclusions by teachers. A classroom assessment may not be well aligned to the district or state curriculum. Moreover, classroom tests, not developed under the high degree of scrutiny that surrounds accountability assessments, may include elements of bias or other unfavorable traits that lead to unreliable, subjective results. Hence, the results from a poorly aligned, subjective classroom assessment tend to be unreliable indicators of a student's preparation for accountability assessments (National Research Council, 2003).

Accountability assessments and classroom assessments are designed for distinct but complementary purposes within an education system. The gaps in each type of assessment appear when either is used beyond its intended role. However, these shortcomings can be overcome when assessments are used together in appropriate, complimentary roles within a coherent assessment system.

Toward a Complete Assessment System

Closing the gap between classroom assessments and accountability assessments necessitates a system in which assessments are designed to support each other. In a recent workshop, a committee of assessment and education experts convened by the National Research Council (2003) identified characteristics of an ideal, complete assessment system. In their consensus, such an assessment system would be *comprehensive*, by including a range of formative and summative

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

approaches to assessments; *coherent,* by including assessments that are compatible with one another and aligned to the same standards, curriculum, and professional development goals; *continuous*, by measuring student achievement over time rather than merely providing stand-alone "snapshots"; and *integrated* into the larger education system that provides teachers with the resources that they need to meet their instructional responsibilities. All of the assessments used in the system should be of a high quality; be a valid, reliable, and objective source of data; and be packaged in a way that is practical for teachers to use (Pellegrino, Chudowsky, and Glaser, 2001). Obtaining these qualities in educational assessments requires some relatively new design approaches, including: ensuring horizontal and vertical alignment, reporting results using a common scale that tracks growth, and producing results that give insights into student learning and misconceptions.

Horizontal and Vertical Alignment

To be comprehensive, coherent, continuous, and integrated, an assessment system must be aligned in two ways identified by Webb (1997): horizontally and vertically (see Figure 1).

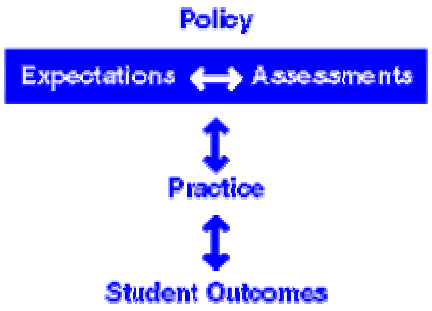

Figure 1. Horizontal and vertical alignment within an education system (Webb, 1997)

Assessments must be *horizontally aligned* to academic content standards and to the other assessments in the assessment system. The process of horizontal alignment makes an assessment system more coherent by causing its content to match the expectations to which students and teachers are held accountable. Moreover, assessments given at sequential grade levels must also reflect the logical order and continuity for each subject area. For an assessment system to reflect the continuous nature of student learning, the content and structure of an assessment should correspond to the logical order in which the subject area is taught. Horizontally aligning assessments puts both their content and the overall structure of the assessment system into this logical order.

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

The assessments must also be *vertically aligned* to other elements of the education system, such as national, state, and local education policy; classroom instructional practices across grades; and student performance expectations. Assessments are frequently developed without fully considering the other elements of the education system. Vertically aligning each assessment ensures that the assessment system is well integrated into the elements of the larger education system. Moreover, vertical alignment contributes to the comprehensiveness of an assessment system by identifying an education system's formative and summative needs that can be filled by classroom and accountability assessments.

Tracking Achievement Growth with a Vertical Scale

For an assessment system to be continuous, assessments should produce results that show achievement growth over time (National Research Council, 2003). One of the most widely accepted methods for reporting student growth is through the development of a "vertical" scale. When linked to the same vertical scale, assessments that are administered at different times produce results that can be compared directly (Jorgensen, 2003; Jorgensen and McBee, 2003). By using a common vertical scale throughout an assessment system, classroom assessments and accountability assessments can track a student's educational progress. Moreover, classroom assessments that use a vertical scale can serve as a reliable indicator of a student's performance on accountability assessments that use the same vertical scale. While vertical scales have long been used in norm-referenced tests, they are now appearing in accountability assessments. Their potential application in classroom assessments is compelling and may help bring continuity to an assessment system.

Providing Insights into Student Learning

Another point of consensus is that the results from assessments should help teachers understand a student's cognitive processes and specific breakdowns in understanding. Such a capability increases the integration of an assessment system into an education system by providing teachers with assessments that directly identify their students' educational needs. Traditional classroom assessments reveal these aspects of a student's learning process, albeit in an unsophisticated manner. However, classroom assessments that incorporate scientific research and professional wisdom about student learning into their results can provide sophisticated insights that are otherwise difficult to attain (National Research Council, 2003; Pellegrino, Chudowsky, and Glaser, 2001). Assessments that include this capability are a recent innovation and are largely still in development.

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

Progress in Assessment Systems Abroad

As the global marketplace continues to become increasingly competitive, nations have been evaluating their education systems to identify opportunities for reform. Before embarking on ambitious research projects, it is worthwhile to examine the progress made in the assessment systems outside the United States, especially those in other English-speaking countries. Not all of the aspects of ideal assessment systems are demonstrated by these examples. However, features of these programs represent practical successes in building coherent assessment systems in which classroom and accountability assessments are designed to function with one another in close alignment to standards and curricula.

ACER Assessment System

The Australian Council for Educational Research (ACER) developed and distributed a system of materials, activities, and classroom assessments to assist teachers in monitoring student progress through levels of achievement in different subject areas. With the goal of helping students reach proficiency as measured by accountability assessments, the ACER assessments were aligned to national content and performance standards. Moreover, resources were devoted to ensuring that the classroom assessments were of a high quality and that teachers were trained to administer the assessments. As the program produced successful outcomes over time, the Australian government adopted the ACER program as its national survey of student literacy to provide policymakers with needed data. In ACER's resulting *Literacy and Numeracy National Assessment* (LANNA), vertical scales are used to report results with comparisons to national proficiency standards for each subject area (ACER, 2004). ACER's efforts demonstrate a successful, practical example of classroom assessments and accountability assessments designed to function together in a coherent assessment system (National Research Council, 2003).

Queensland Assessment System

In 1970, stakeholders in the education system of Queensland, an Australian state, raised concerns that its college entrance exams were too difficult for students to achieve sufficiently high results (National Research Council, 2003). The state responded by developing a system for guiding and monitoring the formative and summative classroom assessments used by teachers to prepare students for the high-stakes exam. In this system, teachers at the local level retain responsibility for developing the curriculum and assessments. State education agencies provide support and guidance to ensure that the assessments are of high quality and are closely aligned with the curriculum. In use for over 30 years, the Queensland assessment system is regarded as successful in helping students to understand and reach achievement expectations on accountability assessments.

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

New Zealand Assessment System

During the 1990s, New Zealand also engaged in a series of reforms of its education system with the aim of increasing student achievement and thereby enhancing New Zealand's global economic competitiveness. Parallel with instituting curricular reforms, the Ministry of Education implemented an assessment system with a wide range of formative and summative assessment tools for classroom teachers. Professional development in assessment was also provided to assist teachers in using these new tools. As New Zealand does not maintain a national accountability testing program, formative classroom assessments have a central role in raising student achievement (Philips, 2000).

A Solution: *Stanford Learning First*™

Despite education research that highlights the significant impact that classroom assessments have on student achievement, significant resources have been expended in the research and development of accountability assessments. Success on accountability assessments clearly depends on the classroom assessments used in daily classroom activities to prepare students. However, to function as a part of a complete assessment system, classroom assessments require more attention to their design and development than in the past. There is a clear need for education researchers, agencies, and test publishers to work with educators in the research and design of classroom assessments that coherently function with accountability assessments (Pellegrino, Chudowsky, and Glaser, 2001).

Harcourt Assessment, Inc. (Harcourt) is currently working toward this goal by developing *Stanford Learning First,*™ an assessment system that incorporates formative and summative classroom assessments that are closely aligned to state content and performance standards. With *ClassViews,*™ the summative component of *Stanford Learning First*, teachers and administrators can identify students who require intervention in advance of high-stakes accountability testing. *ClassLinks*™ provides teachers with a series of formative classroom assessments that give immediate feedback about an individual student's progress in a subject area. Moreover, *Stanford Learning First* analyzes a student's incorrect responses to provide insight into his or her misconceptions about specific subject-area objectives. Specific instructional practices are suggested by this analysis, thereby addressing immediate classroom needs and providing teachers with ongoing professional development. Finally, *Stanford Learning First* offers both paper-based and online administration and scoring simultaneously, a feature identified by education experts as important to the success of future assessment systems, especially in the implementation of formative assessments (Pellegrino, Chudowsky, and Glaser, 2001).

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

With this combination of features, *Stanford Learning First* can be described as comprehensive, by including both summative and formative assessments; coherent and integrated, through its horizontal alignment to standards and vertical alignment to the goals and structure of education systems; and continuous, by using multiple periodic assessments to track student achievement. Stanford Learning First clearly has an important role to play in the development of complete assessment systems.

Conclusion

Developing an ideal, complete assessment system will require significant resources to be devoted to the development of both classroom assessments and accountability assessments. Traditionally, assessment research and development has been directed toward the improvement of accountability assessments. Now, the time is right for the development of valid, reliable classroom assessments to fill the gaps in today's assessment systems (National Research Council, 2003; Pellegrino, Chudowsky, and Glaser, 2001). Carefully designed classroom assessments with innovative features, such as results that use vertical scales and that provide insights into student misconceptions, will empower teachers and students to continuously progress toward meeting achievement expectations. Accountability assessments also deserve renewed consideration for potential improvement and integration within an assessment system.

Stanford Learning First is a significant step toward forming a complete assessment system. By developing this new classroom assessment solution, Harcourt presents a strong answer to the sophisticated, complex needs of the nation's education system. Teachers and students, through the continued collaboration of test publishers, education agencies, and policymakers, will have the tools that they need to meet the achievement expectations and demands of an increasingly competitive global marketplace.

References

Airasian, P. W. (2001). *Classroom assessment* (4th Ed.). Boston: McGraw-Hill.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). S*tandards for educational and psychological testing*. Washington, DC: Author.

Ananda, S. (2003). *Rethinking issues of alignment under* No Child Left Behind. San Francisco: WestEd.

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.

Arter, J. A. (2004). Assessment for learning: Classroom assessment to improve student achievement and well-being. In Wall, J. E. & Walz, G. R (Eds.). *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 463–484). Greensboro, NC: CAPS Press.

Australian Council for Educational Research (ACER). (2004). *Literacy and Numeracy National Assessment (LANNA) brochure*. Camberwell, Australia: Author. Retrieved from http://www.acer.edu.au/tests/school/lanna/documents/LANNA_brochure_2.pdf on September 20, 2004.

Jorgensen, M. A. (2003). *The value of the Stanford scale as a common metric.* San Antonio, TX: Harcourt Assessment, Inc.

Jorgensen, M. A. & McBee, M. (2003). *The new NRT model.* San Antonio, TX: Harcourt Assessment, Inc.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

National Research Council. (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment.* Washington, DC: The National Academy Press.

Nitko, A. J. (2004). *Educational assessments of students*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Philips, D. (2000). Curriculum and assessment policy in New Zealand: Ten years of reforms. *Educational Review 52(2),* pp. 143–154.

Webb, N. L. (1997). Determining alignment of expectations and assessments in mathematics and science education. *NISE Brief 1(2).* Retrieved from http://facstaff.wcer.wisc.edu/normw/1997alignmentbrief.htm on April 23, 2003.

Additional copies of this and related documents are available from:
Harcourt Assessment, Inc.
19500 Bulverde Road
San Antonio, TX 78259
1-800-211-8378
1-877-576-1816 (fax)
http://www.harcourtassessment.com

Copyright © 2004 by Harcourt Assessment, Inc. All rights reserved.