



# Comparing Standards-based Item Banks and Pre-built Tests for Classroom Assessment



Pearson Inc.

August 2005

# Comparing Standards-based Item Banks and Pre-built Tests for Classroom Assessment

## Introduction

The current era of accountability and standards-based reform in education has had many effects on teachers' approaches to instruction in the classroom. Recently, the importance of assessment has been bolstered by the *No Child Left Behind Act* of 2001 (NCLB). With the enactment of NCLB, teachers and students are held accountable to standards in reading, mathematics, and science through the use of annual *accountability assessments*. To monitor student progress toward proficiency measured by these assessments, teachers frequently rely on quizzes and tests administered in the classroom. These tests, used to obtain information about a student's progress against learning objectives while instruction is ongoing, are classified as *formative classroom assessments* (National Research Council, 2003; Pellegrino, Chudowsky, and Glaser, 2001). However, to be useful in the context of the accountability assessments mandated by NCLB, formative classroom assessments should also be based on state standards. Ensuring a match between standards and assessments is no small task and can be overwhelming to teachers who already have a complex, deep range of instructional responsibilities.

This report discusses two solutions which test publishers are proposing for standards-based classroom assessment. The first solution discussed in this report is the use of *item banks* by districts or schools to generate assessments. The second solution discussed in this report is the use of pre-built classroom assessments, an example of which is *Stanford Learning First*,<sup>TM</sup> published by Pearson Inc. (Pearson). Item banks, which have been developed and used in education for many years, have several distinct advantages and disadvantages. By comparison, pre-built classroom assessments provide an immediately useful and beneficial solution for teachers while avoiding many of the drawbacks of item banks.

## Comparing Standards-based Item Banks and Pre-built Tests

### Item Banks

An item bank is essentially a collection of items stored and used to create new assessments at a later date (Anzaldua, 2002; Bloom, Hastings, and Madaus, 1971; Leclercq, 1980; Nakamura, 2001; Rudner, 1998; Smetherham, 1979; Thorndike, 1971). Item banks have been long used in education. Some of the earliest item banks consisted of collections of individual items that were written on cards and subsequently indexed and catalogued (Anzaldua, 2002; Bloom et al., 1971; Leclercq, 1980). A paper-based item banking system can be straightforward to implement but clearly faces limitations in its practical size, scope, and complexity. The current availability of powerful and affordable computers and electronic databases has increased the feasibility of large and sophisticated item banks. In the recent past, a great deal of effort was required to generate a test from an item bank. Today, the processing power of modern computers makes this task more practical (Anzaldua, 2002; Leclercq, 1980; Thorndike, 1971).

### Pre-built Classroom Tests



A relatively new and compelling alternative to item banks is the use of published classroom assessment systems. In this approach, test publishers provide high-quality, pre-built test forms that are designed for educators to use “off the shelf” as classroom assessments. By administering test forms with content that has already been carefully aligned to state standards, teachers can quickly obtain results which indicate a student’s progress toward proficiency in a subject area during the school year. To establish the validity and reliability of the assessments, the test publisher conducts necessary scientific research and carries out the labor-intensive task of matching the individual test items to state standards. The resources available to test publishers enable the production of a high-quality, formative assessment system which teachers can quickly integrate into their daily classroom activities. By using pre-built classroom assessments, educators can return their focus to teaching and improving instruction rather than developing test materials.

This recent development in educational assessment was motivated by the findings of leading education researchers and experts. Traditionally, accountability assessments have been the primary focus of the resources available to test publishers. However, educational research forums and publications have identified standards-based classroom assessments as an essential part of an education system (National Research Council, 2003; Pellegrino et al., 2001). With the emergence of these compelling instructional tools, test publishers have worked to develop practical solutions, such as *Stanford Learning First*.

## Comparing Standards-based Item Banks and Pre-built Tests

### Comparing the Two Solutions

It is instructive to consider these two solutions for classroom assessment and compare their respective advantages and shortcomings. Educators may find that one solution may be more appropriate and suited to their educational needs.

#### Pros and Cons of Item Banking

The most obvious capability of item banks, although one that comes at great expense, is their utility in generating an endless variety of classroom assessments (Nakamura, 2001; Rudner, 1998). These assessments can be customized to an educator's specific needs and can provide tests for whatever purposes are required over the course of the school year, ranging from daily, formative classroom assessments to end-of-course exams. Depending on the design and implementation of the item bank, teachers may also be able to contribute items that they have used, making their items available for other teachers. Each test form generated is also very likely to be unique and therefore more secure, although formative classroom assessments have less of a need for increased security.

Ultimately, the flexibility that teachers have in creating new tests is the primary benefit of an item bank. However, building a usable item bank requires an immense amount of time and resources. Although an item bank is essentially a "large pool of individual questions from which tests can be built" (Smetherham, 1979, p. 57), it is widely acknowledged that "the bank itself is more than just a collection of items" (Anzaldua, 2002, p. 5). Each item must be coded according to the grade-level content standard which it measures (Anzaldua, 2002; Nakamura, 2001). This requirement is especially important if the item bank is going to be used to generate standards-based tests of any kind.

To be capable of generating valid and reliable tests, the item bank must keep track of a broad range of information in addition to the content of each item. Moreover, each test form that is generated should produce comparable results, regardless of the items that were used. For these expectations to be met, certain statistical information known about each item, such as its difficulty and discriminating power, must be tracked. Without this information, the reliability and validity of each test as a measure of student achievement cannot be certain. Hence, each item must be standardized and calibrated through empirical research studies (Nakamura, 2001; Rudner, 1998). Although the methodologies for conducting such studies are well established, a significant amount of effort is required to perform these calculations for today's item banks, which can include thousands of items (Anzaldua, 2002).

## Comparing Standards-based Item Banks and Pre-built Tests

---

In all, these requirements mean that item banks require much more effort to construct than in the past when teachers provided content using material from their own tests (Bloom et al., 1971). Constant maintenance of the item bank (whether performed by the test publisher or by the district), such as the development of new items and the identification and removal of items with undesirable characteristics, may be required for it to remain useful. Considering that “an item bank is only as good as the material that is put into it” (Anzaldúa, 2002, p. 15), each new item that is written for the bank must be reviewed for quality. Moreover, the empirical studies performed on the item bank when it is initially built may have to be repeated to support new items. In summary, educators considering the use of a vendor-supplied item bank should be sure to ask such questions as:

- Who wrote the items, and what training did they receive?
- Were the items field-tested? What kind of data is available for them?
- Are the items aligned to state standards at the appropriate grade level?
- Have the items been reviewed for quality and potential bias?

After the immense effort described above has been completed, a school or district may have an item bank that serves as an important resource for assessment needs. However, the item bank is not useful in the classroom until test forms have been generated. In many cases, it may be up to the classroom teacher to develop the needed test materials. If the tests developed from the item bank are going to be used for formative purposes in advance of state accountability assessments, a number of factors must be considered, including:

- The test blueprint—results will be most useful if the generated test’s blueprint corresponds with the blueprint of the state accountability assessment. Without this correspondence, the results in the classroom may not provide a valid insight into student progress toward proficiency as measured by the accountability assessment.
- Depth of knowledge—besides matching the relevant content standards in a subject area, the items that the teacher selects for the classroom assessment should reflect the depth of knowledge of the standards that are measured by the state accountability assessment.
- Standard weighting—the teacher must construct a classroom test that covers all of the standards but that does not give more or less weight to certain standards than does the state accountability assessment.

## Comparing Standards-based Item Banks and Pre-built Tests

---

- Longitudinal data from the results—teachers may want to measure their students’ progress toward proficiency on the state accountability assessment over the course of the year. When a teacher uses item banks to construct a series of unique forms that are unlikely to be closely related, it may be difficult to interpret results as demonstrating progress.
- Comparability of results—when teachers are constructing unique forms for their own use, administrators may not be able to compare the results from different classrooms reliably, impeding their ability to make resources available accurately where needed. For example, two teachers of the same grade level may each be testing whole-number multiplication. One teacher might only cover multiplication of one-digit and two-digit numbers, while the other teacher covers a representative mix of multiplication problems using different combinations of one-, two-, and perhaps three-digit numbers. Or, the first teacher may pick only the least difficult items to improve student results, while the items selected by the second teacher represent a wide range of difficulty.
- Open-ended (constructed-response) items—if an item bank includes constructed-response items, teachers may score the same items differently, thereby producing unreliable measures of student achievement. Explanations and training in applying the rubrics can reduce this concern for teachers.

The last two points listed above may represent especially significant concerns for educators who are monitoring their students’ progress toward the state accountability assessment. If results from several forms or open-ended questions are not comparable, it can be difficult for teachers to obtain the information that they need to provide differentiated, targeted instruction for each student.

Besides implementation and maintenance, the variety of issues that must be considered for an item bank can exceed its practical value. Moreover, many teachers are unfamiliar with such concepts as test blueprints and the statistical analysis of items. Unless the item bank includes features to assist the teacher with addressing these issues, the work required to produce tests that are truly useful as formative classroom assessments may significantly detract from the teacher’s already limited instructional time.

### Pros and Cons of Pre-built Classroom Tests

As with assessments built using item banks, there are advantages and disadvantages to classroom tests that are built by a test publisher. Clearly, pre-built classroom assessments have less flexibility than item banks. However, the use of a pre-built classroom assessment system can potentially address many of

## Comparing Standards-based Item Banks and Pre-built Tests

---

the issues that arise when an item bank is used as an assessment solution. Most significantly, the school or district is not responsible for developing and maintaining an item bank or generating test forms. Pre-built tests relieve teachers of the challenging task of trying to build tests that follow a blueprint while adhering to essential statistical constraints. In building formative classroom tests, the publisher handles these responsibilities for the teacher through a rigorous assessment development process. This process frequently includes:

- careful alignment of the items to the state’s content standards while taking into account the depth of knowledge and weighting of the standards on the state accountability assessment
- construction of the test forms to reflect both the blueprint and format of the state accountability assessment
- extensive item review to establish fairness and lack of bias in each item used in the assessment
- research studies (such as field-testing of forms) used to gather empirical data that establishes the validity and reliability of results
- standardization based on empirical data from the research studies, enabling results to be compared between classrooms, schools, or districts
- ongoing maintenance and development to improve test items and forms

By following a rigorous development process, test publishers ensure that the assessments produce reliable and valid results for teachers to use. Additionally, test publishers can introduce features that make pre-built tests well-suited for the role of formative classroom assessments. For example, the forms for a particular grade level can be designed to correspond to the typical instructional progression that a teacher uses during the school year. When the forms are built and standardized consistently, scaled scores from multiple administrations can be constructed and used to track a student’s progress toward proficiency. These results can also identify students who require additional instruction in advance of state accountability testing. Moreover, the results can be monitored and reported at the level of the classroom, school, or district, empowering administrators to provide assistance where it is needed.

In developing classroom assessments for teachers, test publishers also have the opportunity to produce tests with innovative capabilities. In addition to freeing teachers from tasks related to developing classroom assessment materials, these pre-built tests can be designed to contribute to and improve instruction in ways

## Comparing Standards-based Item Banks and Pre-built Tests

---

that standard item banks cannot. For example, *Stanford Learning First* items are designed to produce results that indicate a student's attainment of specific subject-area objectives. These results, available as soon as the student has completed the assessment, are delivered through innovative reports that provide the teacher with immediate feedback for planning the next steps in instruction. More significantly, by providing better insight into each student's progress toward achievement, the assessment results enable teachers to differentiate their instruction according to the individual needs of each student in the classroom. *Stanford Learning First* also includes online professional development, which helps teachers to learn to use the results to improve instruction. Hence, the tradeoff of flexibility is more than matched by the innovative benefits which pre-built formative classroom tests bring to educators and students. By using pre-built classroom tests, teachers can focus on improving instruction rather than devoting scarce instructional time to building their own measurement instruments.

### Conclusion

As pre-built classroom tests become more widely adopted by educators, it is unlikely that item banks will cease to be useful. An item bank developed by a third party, such as Pearson's *Insight* item bank, can represent a valuable assessment resource for a school or district. Even when an item bank is accepted, the production of useful classroom tests remains as a central task that often falls upon classroom teachers with already limited time. By comparison, test publishers can build high-quality standards-based classroom assessments that are immediately useful without requiring additional effort by schools, districts, or educators. Because such tests are developed using a rigorous, standardized process and are designed specifically to be used as formative classroom assessments, teachers can rapidly integrate them into instruction. Moreover, formative classroom assessment systems, such as Pearson's *Stanford Learning First*, are bringing innovative approaches to improving and guiding instruction to the classroom in ways that far exceed the capabilities of conventional item banks. Both approaches to classroom assessment have advantages. In considering item banks and pre-built tests, educators must consider the best way to direct resources to meet their students' instructional needs in the twenty-first century.

## Comparing Standards-based Item Banks and Pre-built Tests

### References

- Anzaldua, R. M. (2002, February). *Item banks: Where, why, and how*. Paper presented at the 25<sup>th</sup> annual meeting of the Southwest Educational Research Association, Austin, TX. Retrieved from [http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content\\_storage\\_01/0000000b/80/0d/c8/e2.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/0d/c8/e2.pdf) on May 12, 2005.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.
- Leclercq, D. (1980). Computerised tailored testing: Structured and calibrated item banks for summative and formative evaluation. *European Journal of Education*, 15(3), 251–260.
- Nakamura, Y. (2001). *Rasch measurement and item banking: Theory and practice*. Washington, DC: ERIC Clearinghouse on Languages & Linguistics. Retrieved from [http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content\\_storage\\_01/0000000b/80/0d/94/16.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/0d/94/16.pdf) on May 12, 2005.
- 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.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academy Press.
- Rudner, L. (1998). *Item banking*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. Retrieved from [http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content\\_storage\\_01/0000000b/80/2a/2b/f4.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/2a/2b/f4.pdf) on May 12, 2005.
- Smetherham, D. (1979, February). Banking school knowledge. *British Journal of Educational Studies*, 27(1), 57–68.
- Thorndike, R. L. (Ed.). (1971). *Educational measurement* (2<sup>nd</sup> ed.). Washington, DC: American Council on Education.

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