

R&D Connections

No. 9 • April 2009

What Does It Mean to Repurpose a Test?

By Cathy Wendler and Donald Powers

Should we use a test for purposes other than those for which it was originally intended? Should we give it to groups of people other than those for whom it was originally designed?

There is a substantial market these days for tests of all kinds. There are also a significant number of organizations, such as ETS, that develop and deliver tests. These organizations often compete for common groups of clients—usually government agencies, educational institutions, or businesses—who use test scores to facilitate decisions about individuals (e.g., promotion, graduation, and admission to college) or to determine public policy.

Even when they are on opposite sides of the world, these clients may have similar needs, so it's only natural that a testing organization may strive to serve multiple groups with the same off-the-shelf assessment.

Moreover, it is not difficult to see how an off-the-shelf product can bode well for an organization's finances: New product development is slow and expensive, and when businesses can attract new customers to their existing offerings, the result can have benefits—both for

the organization's bottom line and, in terms of decreased costs, for customers also.

Unfortunately, when it comes to educational testing, there can be a serious downside to this approach: A test's new use may lack the same strong scientific backing as its original use. Given that test scores can carry significant consequences for test takers, this could be a serious concern.

ETS's corporate standards require that the scores we report are fair and meaningful and that the ways in which they are used are defensible.

At ETS, our corporate standards require that the scores we report are fair and meaningful and that the ways in which they are used are defensible. At the same time, our clients need assessments that are developed quickly, are technically sound, and are affordable.

When we repurpose our tests, we ask ourselves:

How do we meet our clients' needs while adhering to the values that have guided us for more than 60 years?

Guidance for Repurposing

In this article, we will use this simple definition of repurposing: using a test either for test takers or for purposes that are different

Editor's note: Cathy Wendler and Donald Powers are, respectively, a senior research director and principal research scientist in the Foundational and Validity Research area of ETS's Research & Development division.

Guidance for Repurposing Tests

These publications may be useful for deciding how to use tests and interpret their scores appropriately:

- *Standards for Educational and Psychological Testing* (American Educational Research Association, National Council on Measurement in Education, & American Psychological Association, 1999).
- *ETS Standards for Quality and Fairness* (Educational Testing Service, 2002).
- *ETS International Principles for Fairness Review of Assessments* (Educational Testing Service, 2007).
- *TC Test Adaptation Guidelines* (International Testing Commission, 2000).

from those for which the test was originally developed.

For this article, we do not consider this definition to include changing the way the test is delivered (e.g., moving the test from paper to computer format), changing the way scores are used in decision making, or modifying a test to make it more accessible to persons with disabilities.

We also do not mean incidental changes in a test's use—naturally occurring changes in the demographics of the test's target population. We recognize, however, that these naturally occurring changes may, over time, make a test less appropriate if it does not measure the knowledge, skills, and abilities of the changed population as well as it measured those of the original population.

So when is it appropriate to repurpose a test?

At one extreme, we could claim that any assessment can be used off-the-shelf, pretty much as is, for any new purpose or group. We argue that this is not really repurposing at all, but merely *relabeling*, akin to putting old wine into new bottles and giving it a new name.

At the other extreme, we could claim that every assessment must be treated as if it were brand new if it is to be used for any new purpose or group. Under this view, any existing evidence to support the meaningfulness and fairness of test score inferences would be considered inappropriate or irrelevant.

Neither extreme is realistic. In this article, we discuss the issues that testing organizations should consider when seeking the appropriate middle ground.

Where can testing organizations look for this kind of guidance? Several sets of standards and guidelines provide a clear mandate for evidence to support new uses of a test.

First, the publication *Standards for Educational and Psychological Testing* provides explicit criteria for evaluating tests and testing practices. These joint standards—so called because they are backed by three professional organizations with a strong interest in measurement—represent current consensus among professionals in education, psychology, credentialing, and other areas regarding appropriate testing practices.

The authors of the joint standards—the American Edu-

cational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME)—consider it to be a “professional imperative” (1999, p.viii) for their members to observe them.

At ETS, staff members also consult the *ETS Standards for Quality and Fairness* (Educational Testing Service, 2002).¹ The ETS standards are an extension of the joint standards, but have been “...tailored to ETS’s specific needs and circumstances” (p.2). These guidelines, which are no less rigorous than the *Standards for Educational and Psychological Testing*, provide a clear means for evaluating the quality of assessments.

The intention of the ETS standards is to provide guidance, but not to “...stifle adaptation to appropriate new environments” (p. 1), which is important for repurposing efforts.

Also relevant are the guidelines in the *ETS International Principles for Fairness Review of Assessments* (Educational Testing Service, 2007), which ensure that assessments used in other countries are fair and appropriate for the cultures of these countries. The guidance offered by these principles is especially pertinent when repurposing tests for new international markets.

Finally, the International Test Commission (ITC), a committee representing a number of international groups, has developed guidelines for adapting psychological and educational tests for use in various linguistic and cultural contexts (International Testing Commission, 2000). According to these guidelines, it is

incumbent on test publishers to identify and remove any aspects of test questions that might hinder international test takers from fully demonstrating their knowledge and skills.

Clearly, these guidelines show that repurposing a test requires gathering new evidence. The difficulty, however, is in knowing how much and what type of evidence is needed.

A Moderate View

ETS takes a moderate view on repurposing tests. This view acknowledges that it is wasteful not to take advantage of the good work carried out to support the original development of an assessment, including carefully developed content specifications for the test and the writing, review, tryout, and revision of test questions.

At the very least, however, new clients should thoroughly review the content that the existing test covers and either endorse the test in its current form or offer suggestions for its modification.

Even more to the point, experts who best know the new test-taking population or the needs of the new score users should review the fairness and relevance of the existing set of test questions. Such evaluations of the existing test may provide some clues as to how it will function in a new setting, but this is not a substitute for actually trying out the test with the new group of test takers.

Previous studies showing that the test’s scores are fair and meaningful may provide some support for the new use of a test, but their major value is in helping to design new studies that include the new group of test takers or address the new use.

The standards and guidelines are clear in their mandate for evidence to support new uses.

¹ http://www.ets.org/Media/About_ETTS/pdf/standards.pdf

Even if the testing program has expended great efforts to support the test's original purpose, more work may be needed to support the claims that might be made about the test's outcomes in its new context.

How much additional work is required depends on how similar the test's proposed new use and test takers are to the ones for which the test was originally designed. Three examples follow:

Let's say a U.S. college decided to adopt an existing admissions test—one that it had not used in the past. There is plenty of information to suggest that the leading tests used for this purpose in the United States are appropriate for a wide variety of U.S. institutions, since many studies have documented the appropriate use of these tests in this type of decision process. As a result, we would need to do relatively little work to support this new use of the assessment.

However, using the same admissions test for college admissions internationally requires more diligence, as the requirements of international institutions and the students they serve may differ significantly from those of U.S. institutions. Some validity research is required in order to determine whether the test can provide meaningful scores in its new context.

Finally, using the admissions test, or even questions from the test, to screen job applicants in another country is both a very different purpose and a very different group of test takers. In cases like this, one needs to proceed much more cautiously. A considerable amount of research is necessary to determine whether it is appropriate to use the test or test questions in this way.

One methodology that is sometimes relevant for repurposing is that of *validity generalization*—applying evidence gathered in one situation to other similar situations. This approach is endorsed in the standards as one way to

establish scientific support for a test in a different, but similar context.

There is now considerable evidence that suggests that a test is likely to be appropriate across a number of situations, if the situations are reasonably similar. But as the context in which the test will be used becomes less like the context (and the group of test takers become less like the original group of test takers) for which the test was intended, this methodology offers less guidance.

A Typical Scenario

What usually happens when a test is considered for repurposing? Although there are certainly a variety of scenarios, one that commonly occurs at ETS is the following:

A client approaches ETS with a need to assess a group of test takers that, at least on the surface, seems similar to one that we currently serve. We or the client identify an available test that seems to contain content and measure skills or knowledge that are of interest to the new customer—who, however, does not have the time to conduct a thorough review of the test content to confirm whether it actually corresponds to the perceived needs.

The customer typically wants the test soon and needs it to provide high-quality, accurate results. In addition, compared with ETS's more traditional customers, the new customer may not be as familiar with test development procedures.

Even with these difficulties, however, there is usually much to build upon when we set out to repurpose a test.

Some of the most valuable of the available resources are the existing test questions themselves. In fact, in our experience, much of the discussion (and activity) regarding repurposing occurs not at the level of the test, but rather

Validity First

The guidelines in the ETS Standards for Quality and Fairness call for the following considerations when a test is repurposed for a new use:

1. Each proposed new use of a test must fit within the ETS mission – to advance quality and equity in education by providing fair and valid assessments worldwide.
2. It is imperative to state what is being measured (and for what reason), for whom the test is intended, and what will be done with the information that the test provides.
3. The types of evidence that can support the claims about what the test measures need to be determined.
4. It is important to demonstrate the extent to which the test scores live up to the claims made about their use and interpretation.
5. Any unexpected negative consequences of using the test need to be explored.
6. ETS should work with test users so that ultimately they know how to gather evidence on their own so that over time they can make a stronger and stronger case for using the test.

at the test question level. This makes sense in that test questions are concrete and tangible in ways that content specifications are not.

Moreover, good test questions are not something to be discarded after a single use, but rather something to be treasured and reused, if possible. Thus, it is perfectly natural to want to reassemble, repack, and recombine existing test questions.

However, if test questions are merely “stitched” together from various sources without a plan for how they should work together as a test, the result is typically a group of high-quality questions backed by (good) data gathered from a group of test takers that is not representative of the test takers who would take the repurposed test.

ETS’s standards require a plan for assembling test items with a clear purpose, a model for generating test scores, a set of clearly defined claims for the meaning of test scores, and a plan for validating the inferences made from test scores.

A Focus on Validity

At the heart of each attempt to repurpose an ETS test is the concern for providing scores that are fair, meaningful, and defensible. That is, the repurposed test should provide scores whose meaning can be interpreted with a high degree of confidence. As defined in the joint standards, “Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, NCME, & APA, 1999, p. 9). Focusing on validity provides the pathway to ensure that fair, meaningful, and defensible scores are created.

What does that mean at ETS?

First, each proposed new use of a test must fit within the ETS mission—to advance quality and equity in education by providing fair and valid assessments worldwide. In most cases, the stated intentions of a new client suggest how well they conform to the ETS mission. This fit is considered explicitly at the outset of any proposed repurposing.

What to Do When Repurposing a Test

These steps can help to make an existing test's scores fair, meaningful, and defensible for a new purpose or new group of test takers:

1. Identify the differences between the test's original use and its proposed new use.
2. Develop a plausible argument as to why the test should function as expected with the new test takers.
3. Identify the types of evidence that must be obtained, both short- and long-term, to support the intended use and interpretation of the test's scores.
4. Create a plan for dealing with the unexpected and for modifying the test in an expeditious manner, if ultimately required.

Second, it is imperative to state what is being measured (and for what reason), for whom the test is intended, and what will be done with the information the test provides. Repurposing a test requires attention to two key aspects of validity:

- *The degree to which the existing test appropriately measures the dimensions considered important for the repurposed use*—For instance, if a test is intended to predict success in traditional academic settings, then its scores should measure the most important academic skills.
- *Factors that influence test performance outside of the skills or ability being tested*—These need to be identified since they may have unknown impacts on results from the new group of test takers. For example, if the test is not intended to measure verbal ability, then the reading level required to answer test questions may be of concern if the repurposed assessment is to be given to a group of test takers who are less proficient in English than in the group for whom the test was designed.

Third, it is important to determine the types of evidence that can support the claims about what the test measures. This evidence can take a variety of forms: relationships of test scores to scores from similar tests, test-taker improvement over time, comparisons between the test performance of different groups of test takers, and information about the strategies that test takers use to answer test questions, for example.

Fourth, it is important to demonstrate the extent to which the test scores live up to the claims made about their use and interpretation. Doing this means designing studies that answer the most important claims that clients want to make about a test: Does it predict subsequent college performance? Does it allow test takers to demonstrate the knowledge and skills they need for certification in a profession? Does it facilitate appropriate student placement into courses? Test users are cautioned against using the test for purposes that are tempting but not supported by empirical evidence.

Fifth, any unexpected negative consequences of using the test need to be explored. This means, for example, investigating differences in the performance of demographic groups when these differences seem to be related to gender, native

language, country of origin, race or ethnicity, socioeconomic status, or other factors that may signal unfairness.

Sixth, ETS works with test users so that ultimately they know how to gather evidence on their own so that over time they can make a stronger and stronger case for using the test.

And finally, above all else, both the joint standards and the ETS standards mandate that, if important factors change, we must reexamine the existing evidence supporting a test's use. If necessary, new evidence must be gathered. This is key to supporting any effort to repurpose a test for a new use.

Steps for Repurposing

Above, we've described some principles we would apply at ETS when considering whether to repurpose a test. For all testing organizations, however, the processes involved in repurposing a test should be similar.

- *Identify explicitly the differences between the test's original use and its contemplated new use.* Examine the information that has been used to support the original use of the assessment and identify areas in which the existing evidence does not support the new use. Ask: "Is the evidence that supports the initial purpose of the test also sufficient for meeting its new purpose? If not, how does it fall short?"
- *Develop a plausible argument as to why the test should function as expected with the new test takers.* Then, try out the test on a sample of the new test takers to determine, at the most basic level, if the test questions work for them and if the test as a whole functions adequately.
- *Focus on validity.* Identify the types of evidence that must be obtained, both short- and long-term, to support the intended use and interpretation of the test's scores. Create a plan—and realize that it is not reasonable to expect that all of the necessary evidence can be gathered quickly or in a single study.
- *Finally, create a plan for dealing with the unexpected and for modifying the test in an expeditious manner, if required.*

In summary, it is likely that many existing assessments can help meet the considerable demand for assessments of all kinds. However, there are steps that can (and should) be taken to gather appropriate evidence to support the new use of an existing test. Ultimately, these steps should ensure that any repurposed assessment meets professional standards for testing and thus provides test scores that are fair, meaningful, and defensible.

References

- American Educational Research Association, National Council on Measurement in Education, & American Psychological Association. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton, NJ: Author.
- Educational Testing Service. (2007). *ETS international principles for fairness review of assessments*. Princeton, NJ: Author.
- International Test Commission. (2000). *TC test adaptation guidelines*. Retrieved March 30, 2009, from <http://www.intestcom.org/Guidelines/test+adaptation.php>



Listening. Learning. Leading.®

R&D Connections is published by

ETS Research & Development
Educational Testing Service
Rosedale Road, 19-T
Princeton, NJ 08541-0001

Send comments about this publication to the above address or via the
Web at: <http://www.ets.org/research/contact.html>