Practical Advice on Proficiency Testing

Peter Behuniak, Jr.

Proficiency testing is a fact, and educators should take the lead in developing tests and setting standards carefully.

Minimum competency or proficiency testing has leapt from relative obscurity a few years ago to a position of central importance in American education. It is no longer pertinent to argue the pros and cons; three fourths of the states have mandated some form of testing, so educators are now concerned primarily with implementation. Early evidence supports the potential value that proficiency testing can have, but the sudden sweeping changes that must be made by millions of people in thousands of school districts suggest that it may not succeed. I will discuss some of the issues that need to be considered at this time, and offer practical recommendations about how to proceed.

The Clouds

The first and perhaps most significant problem is that many local districts are under pressure to act right now. In most places legislators who require educators to institute proficiency testing programs conveniently allow state and local education officials the right to determine how to implement them. (This is similar to the ancient practice of forcing slaves into the arena to fight the lions. The decision about whether to fight was made by government leaders,

1 The subtle distinctions between proficiency testing and minimum competency testing will not be developed here because of space limitations and because most educators use the terms interchangeably.


but the decision about how to fight was generously considered to be the slave's problem.) Nevertheless, I believe educators can do a respectable job of laying the foundation for a solid program in spite of the atmosphere of coercion that is so prevalent. To do this, local teachers and administrators must apply common sense to the plans laid out for them. Being aware of the problems discussed in this section can help, but there is no substitute for experience. Educators can and should question procedures that their experience indicates as being unsound.

This may seem like an overly general charge, and it could be interpreted as an argument against any change, but I think most educators will understand the wisdom of taking slow and deliberate action. When the current furor over falling standardized test scores subsides, teachers and administrators will have to accept responsibility for the newly implemented programs in our schools. This is as it should be; as it must be. It is simply not responsible behavior to stand by meekly while legislators (or anyone else) force underdeveloped and untested proficiency testing programs on schools, then later to disclaim responsibility for inadequate programs under the pretext of not having wanted them anyway. If educators are to be effective in guiding the sensible use of proficiency testing, they must be familiar with potential problems.

Proficiency testing programs can be no better than the tests themselves. Consider a common progression of events when developing a proficiency test: 1. Decide to have a test. 2. Create (or select) an objective-referenced test in the appropriate content area. 3. Set a standard or minimum passing score. 4. Administer the test. 5. Interpret the results. Since objective-referenced, multiple-choice tests are the form most commonly used in proficiency testing situations, it is critical that basic test development procedures (for examining item clarity, difficulty, reliability, validity, and so on) be followed in step two above. If this is not done, subsequent activities will be nothing more than futile exercises. Even worse is the likelihood that poor attention to detail at the start will obscure future efforts to develop and refine the test. As a simple example, the lack of pilot testing to establish test reliability can lead to changing percentages of students passing and failing from year to year with no relation to the efforts and abilities of the teachers and the students involved. If this situation occurs, the officials involved will probably not attempt to correct past errors but will instead simply misinterpret the results.

Consider a few of the events taking place in the Florida competency program. In October 1977, competency tests in mathematics and communication skills indicated failure rates of 36 percent and 8 percent respectively. After $26.5 million was spent on remediation, retesting in October of 1978 revealed failure rates of 26 percent and 3 percent in the same content areas. State Education Commissioner Turlington announced that "the gains prove that students can achieve" because "the final scores of the 1978 administration speak for themselves." That conclusion is unsound because two basic principles of test development and administration were violated. Parts of both tests were changed with no assurance that the new parts of the 1978 test would be comparable to the 1977 test. Further, retesting some of the same students with the same items one year later (the majority of items on the tests were unchanged) allows for some gains solely because of familiarity with test items.

Another problem that must be faced by test enthusiasts is that there is no single best method for setting a standard, or minimally acceptable level of performance, for a proficiency test. With the number of suggested methods being over two dozen and climbing, how should local educators select the method they will use? Or should they use several methods? And then, should they average the results, or choose the most acceptable one? How high or low should the resulting standard be, given a particular method?

Unfortunately, direct answers to these questions are not currently available. Although a few careful comparisons among methods have been made, and others are in process, not enough is known to provide definitive answers. Nor will these issues be resolved anytime soon; members of the measurement community have just recently turned their attention to these issues (despite the existence for many years of methods such as Nedelsky, 1954), and longitudinal studies are needed to answer some of the questions. For example, questions about the impact of proficiency testing on subsequent student achievement cannot be answered without studies of several years' duration. No amount of wishing, hoping, or legislat-
ing is going to change this. That is precious little comfort to teachers and administrators faced with imminent deadlines, but awareness of these problems may give them a realistic understanding of the difficulties they face. At the very least, this discussion should explain why all standards, regardless of how they are generated, must be viewed as tentative and subject to review and revision.

The final cloud on the proficiency test horizon is that perennial favorite—inadequate resources. The reason this problem deserves special note (when other educational programs are also inadequately supported) is that cutting corners in proficiency testing is so easy that some proponents could be shortchanging their programs without knowing it. Take the issue of setting standards, for example. One district might choose an entirely arbitrary level—say 70 percent—as the standard for all tests. Another district might help students scoring lowest and moving to higher difficulty among individual tests and the second approach circumvents the purpose of proficiency testing: to determine whether particular students are minimally proficient. In the latter case, it would be far better to use existing diagnostic tests.

The Silver Linings

Proficiency testing of basic skills offers a practical method of shifting attention to how well students do on tests based on locally developed objectives, in contrast to our long standing dependence on norm-referenced standardized test scores. This shift can have more than one positive effect. It provides a tool for assessing student progress without setting up an atmosphere of competition (where the quality of a student’s score is measured by the percentage of students he/she surpasses). Applied intelligently, proficiency testing can succeed in replacing norm-referenced tests, which simple objective-referenced testing has failed to do. This is true because proficiency testing provides a critical element that most objective-referenced tests do not—a standard or yardstick against which to measure the quality of a student’s work. With this helpful ingredient, educators and parents who wish to avoid fueling classroom competition can turn to an empirically sound alternative.

Another advantage of proficiency testing is that the underlying principles are in keeping with recommended practice even if the terminology and some of the problems are new. The process of setting well-defined goals (for teachers and students) and measuring progress toward those goals is an admirable activity already in use in many places. So, while proficiency testing goes beyond these activities (by requiring performance standards and prescribing outcomes based on students’ scores), it can coexist in a mutually supportive arrangement with current educational practice. In fact, I believe a good proficiency testing program can reduce the volume of classroom testing that plagues teachers and students today.

An advantage that will not be apparent to those unfamiliar with proficiency testing lies in the beneficial effects derived from the process of teachers and parents who wish to avoid fueling classroom competition can turn to an empirically sound alternative.
administrators (and parents and students) working together to set standards. It is probably obvious that some benefit would be gained through the thoughtful, cooperative efforts of the concerned parties; that each group would tend to see things from a slightly different perspective, and that the sharing of these ideas would be a positive factor. What may not be obvious is the positive effect of the standard-setting process on the tests and on the individuals involved. Those participating in a standard-setting session are generally required to think about the test items as would a student taking the exam. Doing so helps them identify problems of wording, logic, misspellings, and ambiguities missed earlier. The resulting improvement in the tests helps reduce student frustration and alienation and yields more accurate results. The direct benefits to the individuals involved arise from their increased awareness of the curriculum being addressed by the test. They also develop a more realistic understanding of the limitations of paper and pencil tests. I have witnessed teachers who, apparently for the first time, discovered just how unfair to students a poor test item can be (usually because that teacher could not decipher the question properly). This leads to a greater appreciation on the part of teachers (and administrators and parents) of the difficulties faced by students taking the test. Even better is the tendency for this awareness to spread to other staff members who did not have the opportunity to join the sessions. The net effect, then, may be to enhance staff and community relations and to broaden both groups’ understanding of the testing process.

The preceding discussion would be meaningless if it were offered in a vacuum, out of context. The context in which these points must be viewed is the current state of proficiency testing: the state of the art and the state of the legislation. A look around shows clearly that legislation is surpassing the current state of the art at nearly every turn. The necessary question then becomes, "What can local districts do if they are being forced to comply with state legislation?" While the following suggestions are given with this question in mind, they may be helpful to any district considering implementation of a proficiency testing program.

Recommendations

Initial planning. If your state is assuming total responsibility for developing, administering, and interpreting the tests, then much of what follows is superfluous. If, however, your district retains at least a vestige of power to control what takes place in its schools, these considerations should prove helpful. View the decision to develop a proficiency test as the first step of an important but extended journey. It is simply not practical to expect a one-shot effort to be sufficient. The staff should know that they have before them the tasks of creating the test, setting the standard, administering the test, examining the scores, refining the standard, and repeating the processes of administering, examining, and refining again and again. Even if much more were known about the implications of setting standards and using test results, a commitment to the refinement process would still be necessary in order to make adjustments for changes in
LEARNING DISABILITIES
The Struggle from Adolescence
Toward Adulthood
William M. Cruickshank
William C. Morse
Jeannie K. Johns

This book deals directly and exclusively with the nature and needs of the adolescent with learning disabilities. The first part relates five specific case histories. The second part discusses teaching in the secondary school. The authors present a core educational regimen which will prepare children for life and will insure the integrity of the youth's individual personality. They set forth a teaching method for establishing and testing student performance in the key areas of reading, organization, study skills and written language.

A comprehensive and far-reaching project representing years of research and testing, Learning Disabilities will be of great value to those charged with implementing programs in curriculum and instruction on the secondary school level.

304 pages, 15 figures, tables, references
Cloth $18.00
Paper $ 8.95

From your local bookseller or prepaid from:
SYRACUSE UNIVERSITY PRESS
Dept. S
1011 East Water Street
Syracuse, New York 13210

Addressing these issues early greatly increases the chance of developing a successful program. This is accomplished by eliminating misconceptions and unreasonable expectations, thus avoiding the disillusionment that occurs when expectations are not met. It is not necessary to definitively answer all of these questions. What is important is that these topics are discussed openly and that differences of opinion are brought to the surface. It is certainly preferable to raise questions, even though answers may have to be deferred, than to leave people with the misconception that their unique solutions are universally accepted.

Purpose. If there is any flexibility in the matter, resist tying proficiency test results to promotion or graduation, particularly on a first administration. Measurement experts know about the many things that can go wrong, factors that contribute to the error in test scores. While error cannot be eliminated completely, it can be reduced to within tolerable limits— but only if the test is examined and refined. No one can fire off a perfect 50-item test on the first try. By using the test for only remedial purposes at first, the pressures and problems that can arise will be lessened.

Educators who have been handcuffed by legislation mandating the link between proficiency tests and graduation and/or promotion can attempt to postpone this requirement by requesting that the first one or two administrations be viewed as trials. This is not only more just for students since it provides up coming classes with greater warning and a more fully developed test, but it allows time to iron out problems that develop. Further, the task of adjusting and correcting the test is made easier because the stakes are not as high when decisions involve remediation rather than promotion or graduation.

Standard setting. There are several points that need to be made about the process of setting standards. The first is that these suggestions should be considered tentative, pending further research and experience. Although they represent my own preferences and experiences, they are generally consistent...
with the shared experiences of leading practitioners in the field.  

Involving as many people as possible in the process. Factors such as the content and grade level will dictate whether student involvement is practical. The time of day at which meetings are held may limit parent involvement, but there are always compromise methods to accomplish some involvement where there might have been none, or more involvement where there might have been little. The support for the program that will be gained will be well worth the effort.

Though many methods of standard setting have been proposed, most districts should find three methods to be sufficient. The Angoff, Nedelsky, and Ebel methods have been employed most frequently and have the most known about them. These methods, particularly the Angoff and the Nedelsky, facilitate identification of poor items by guiding the standard setters through an inspection of the test on an item-by-item basis. These methods are recommended because they are direct, relatively easy to apply, and can be utilized in school districts without a great deal of expense or outside involvement. Specific guidelines for applying each of these methods are too extensive to be given here but can be found in other sources.

Although each of these methods is independent of the others, it is advisable to apply at least two, if not all three, of the procedures. The application of more than one method to the same test allows for comparisons among the standards generated by the methods. While some would view the likely discrepancy among standards as evidence for scrapping the whole process, I believe that an examination of the differences among standards is necessary and enlightening. It helps the participants shed whatever desire they may have to accept a standard as absolute. It also sets the stage for attempting to refine the standard by demonstrating visibly that an acceptable standard can be reached only through careful thought and dialog over an extended period of time. The need for a long term commitment is once again made clear.

A final point about the standard-setting process concerns the amount of time required to generate a standard. These requirements vary from method to method and from test to test, depending on such factors as length and subject matter. Generally, however, one or two working days should be sufficient to complete the group activities required. One of these group activities that can consume a great deal of time, however, is the definition of what constitutes minimally acceptable performance. Although space does not permit a thorough discussion of all the aspects of reaching such a definition, previously cited references are available and I will, therefore, offer only the following comments. The main reason arriving at a mutually agreeable definition of minimum performance can be so difficult is the diversity of opinions about how to operationalize the broad concepts identified by the district's goals and objectives. There is no argument that high school graduates should be able to read, but there is a great deal of disagreement about what they should be able to read, and how well they should be able to read it. If a school district is to develop a workable proficiency testing program, the individuals participating in the standard-setting process should recognize that these differences of opinion are natural and must be accepted. The open discussion of the specific skills that should be expected of a minimally proficient student is critical to the intelligent development and implementation of a proficiency test. It is not practical or desirable to attempt to force consensus from the group. It is important, however, to allow enough time for discussion so that all points of view can be heard and responded to by other group members. It may also be wise to hold this discussion several days in advance of the actual standard-setting sessions. This allows ideas to sink in and, if necessary, permits time for further discussion.

The course that proficiency testing takes in the public schools is up to us to decide. Like any innovation, it requires resources in the form of time, money, and effort. And, as with any innovation, the expenditure of resources will not guarantee success. Nevertheless, we can increase our chances of success as we guide our schools into uncharted waters by proceeding cautiously, relying on the best available information and common sense to plan our course, and recognizing that new developments are likely to necessitate changes in heading as we grow through experience.
