



PROJECT MUSE®

Preparing for High-Stakes Testing

Cengiz Gulek

Theory Into Practice, Volume 42, Number 1, Winter 2003, pp. 42-50
(Article)

Published by Ohio State University College of Education



➔ For additional information about this article

<https://muse.jhu.edu/article/41411>

Preparing for High-Stakes Testing

The passage of the No Child Left Behind Act has spotlighted testing and accountability in U.S. public schools. This federal statute calls for a dramatic expansion of state-level high-stakes testing. Educators need to prepare students for these tests in ways that do not detract from real learning. In addition, school practitioners must become assessment literate in order to make the maximum use of test results. This article addresses appropriate and inappropriate test preparation practices, as well as some practical aspects for becoming assessment literate.

STATE-MANDATED TESTING PROGRAMS have become more prevalent than ever. As noted in *Education Week* ("Quality Counts," 2002), every U.S. state has some form of testing program. Until the enactment of the No Child Left Behind Act, states had some leeway in determining whether to attach high stakes to the test results. However, with the passage of this act, every state-mandated testing program has become high stakes for schools and districts. One outcome of such an emphasis on high-stakes tests is increased pressure on public school practitioners to raise scores. These pressures can lead to a distorting effect on teaching and learning.

Cengiz Gulek is the director of assessment and evaluation at the Pleasanton Unified School District, CA.

This article discusses how to prepare students for high-stakes tests in ways that do not detract from real learning. Also discussed is the need for teachers to take charge of the assessment process by becoming assessment literate. This does not mean improving one's ability to "teach to the test." Rather, it means becoming informed as to what constitutes appropriate and inappropriate uses of test results. In addition, it means staying apprised of the latest research on how students learn, and how best to assess what they know—two topics discussed by Chudowsky and Pellegrino (this issue).

Ways to Prepare Students for High-Stakes Tests

Adequate and appropriate test preparation plays an important role in helping students demonstrate their knowledge and skills in high-stakes testing situations. Norton and Park (1996) found a significant relationship between test preparation and academic performance. Chittooran and Miles (2001) also concluded that adequate test preparation significantly improves student attitudes toward test taking and, hence, actual performance on high-stakes tests. In a meta-analytic study of test preparation practices, Miyasaka (2000) identified five types of test preparation practices that help students more fully demonstrate their knowledge and skills on high-stakes tests. These include (a) teaching the content domain, (b) using a variety of assessment approaches and

formats, (c) teaching time management skills, (d) fostering student motivation, and (e) reducing test anxiety. Each of these is discussed below.

Teaching the content domain

When high-stakes are attached to the test results, there is a tendency for educators to emphasize the objectives from the content domain that are sampled on the test. More often, instruction becomes limited to the content areas that teachers know will be tested, and other content areas are neglected. A teacher should not engage in instruction that addresses only those portions of knowledge included on the test. While this may raise scores, it will not build students' knowledge and skills in the broader subject area. As suggested by the examples given in Table 1, appropriate test preparation practices produce student learning that is robust (i.e., generalizable to contexts outside of performance on the test). Inappropriate test preparation practices focus only on raising scores on the test. In order to give students a fair chance to demonstrate what they know and can do, it is essential to expose them to all curriculum objectives to be mastered at their grade level. When this is done, test scores will most likely take care of themselves.

Using a variety of assessment approaches and formats

A number of states are moving toward using a mixture of item formats on their tests, although

the multiple-choice mode still dominates. *Education Week* ("Quality Counts," 2002) noted that 47 states use both multiple-choice and open-ended questions on their tests. Students are expected to demonstrate their knowledge and skills by responding to a variety of open-ended question formats, including short- and extended-response questions and essay prompts. For example, in addition to multiple-choice questions, the California High School Exit Examination (CAHSEE) has several writing essays that make up 30% of the English-Language Arts score. To help students do their best on the CAHSEE, teachers in California need to use these question formats in class as part of their test preparation. Apart from preparation for tests, it is good educational practice to expose students to a variety of assessment approaches and formats because this allows them to apply their knowledge and skills in multiple learning situations. In addition, because students have different learning styles, using multiple modes of assessment gives the teacher more opportunities to see what students know and can do, and to adjust instruction accordingly.

Teaching students time management skills

One thing that can hinder students' ability to perform well on a high-stakes test is a lack of time management skills. Time management skills are even more critical for students with special needs. For example, Jakupcak and Rushton (1992) found that when teachers in their study focused on teaching

Table 1
Appropriate and Inappropriate Test Preparation Practices

Appropriate Test Preparation	Inappropriate Test Preparation
Teaching the content of the domain to which the user wishes to infer	Engaging in instruction that limits one's ability to infer from the test score to the domain of knowledge/skill/ability
Teaching test-taking skills	Limiting content instruction to a particular item format
Teaching toward test objectives if the test objectives match the domain objectives	Teaching only those objectives from the domain that are sampled on the test
Ensuring that students understand the test vocabulary	Using an instructional guide that reviews questions from the latest issue of the test
Assessing students on various aspects of the content domain	Limiting instruction to the actual test questions

Source: Mehrens (1991, April)

time management skills, all students, including those with disabilities, demonstrated proficiency on the course exams. Clovis (1999) suggests that the most direct and simple way to build time management skills is to give a few tests with time limits throughout the year so that when faced with a timed test, students do not panic. Time management skills in other areas (e.g., studying for tests) can also produce better performance on the test. For example, Loulou (1997) suggests three ways to help students study for a high-stakes test:

1. Daily Reviews – Have students conduct short reviews of lecture notes before and after class. Recommend that they begin reviewing after the first day of class.
2. Weekly Reviews – Suggest students dedicate about one hour per subject to reviewing assigned reading and lecture notes.
3. Major Reviews – Have students start the week before an exam and study the most difficult subjects when they are the most alert. Suggest that they study 2-5 hours with sufficient breaks. When possible, have them review their answers to tests given during the year so they can see where they made mistakes.

Time management is not only a skill needed for performing well on high-stakes tests; it is also a skill that a student can use in every stage of life. Teaching students time management skills will increase their chances of performing well in high-stakes testing situations.

Fostering student motivation

A positive outlook about high-stakes tests is nearly as important as one’s knowledge of the content area that will be tested. Attitude plays a significant role in student performance, particularly at the lower grades. White (1989) found that there was a consistent relationship between a positive attitude and higher levels of reading achievement in grades 1 through 8. Also, Roderick and Engel (2001) concluded that students with high levels of work effort generally made greater-than-average learning gains. Similarly, Roth and Paris (1991) concluded that student motivation was a significant factor in doing well on standardized tests.

In their study of high school students’ science achievement, Haydel and Roeser (2002) identified three motivation patterns that affect student engagement and achievement. Table 2 summarizes

Table 2
Motivation Patterns, Purpose of Engagement, and Learner Characteristics

Motivation Pattern	Purpose of Engagement	Characteristics of Learners
Intrinsic-Mastery	Improving their skills, competencies, and intelligence	Enjoy learning; seek out challenges; persist during difficulties; use adaptive problem-solving and learning strategies; and show continuing motivation to learn in a subject domain outside of formal learning settings
Ego-Success	Proving of one’s fixed ability or the hiding of one’s fixed inability	View achievement situations as opportunities not necessarily to improve skills and competencies, but rather to prove superior relative ability; have high confidence in their abilities; show negative affect during learning and the use of nonoptimal learning strategies
Helpless	The pursuit of goals in which proving ability or hiding inability relative to others is a central aim	Lack confidence in their abilities and are occupied with the goal of hiding their perceived sense of incompetence; have increased negative affect in achievement settings; avoid challenge seeking; fail to pursue tasks in the face of challenge; and have performance deficits

Source: Haydel & Roeser (2002)

these patterns, as well as the purpose of engagement and learner characteristics that accompany each.

Haydel and Roeser (2002) found that girls were more likely to exhibit the helpless pattern and boys were overrepresented in the ego-success motivation pattern. However, gender was not a significant predictor of perceptions of efficacy. It was further concluded that adaptive motivational patterns, such as intrinsic-mastery and ego-success, were associated with more positive perceptions of multiple-choice and constructed response tests. There was no significant difference among the three motivational patterns regarding students' perceived efficacy for performance assessments. Findings from this study suggest that school practitioners need to understand the motivation patterns of students and use this understanding to help students prepare for the types of high-stakes tests they are required to take. How can this be done? Guthrie and Wigfield (2000) suggest five factors involved in increasing motivation through classroom instruction:

1. Learning and knowledge goals – setting core learning goals that are codeveloped by the teacher and the students;
2. Real-world interactions – making connections between the academic curriculum and the personal experiences of learners;
3. Interesting subject content – students will devote effort, attention, and persistence to topics that are enjoyable and intriguing;
4. Strategy instruction – providing direct instruction, scaffolding, and guided practice. Development of intrinsic motivation is strongly dependent on students' competence; and
5. Praise and rewards – giving informative compliments that make learners feel a sense of accomplishment and pride in their work.

Reducing test anxiety

With the greater emphasis on high-stakes tests, test anxiety is becoming a troublesome and common condition among students in public schools. Research indicates that test anxiety may exert a debilitating effect on student performance. The higher the anxiety level, the lower student performance tends to be (Berliner & Casanova, 1988; Hancock, 2001; Smith, Arnkoff, & Wright, 1990). Gender and racial differences may also play an

important role. Crocker, Schmitt, and Tang (1988) found that a combination of race, gender, and test anxiety factors influence student performance on standardized tests. Female non-White students seem to have higher levels of test anxiety than males and White students in standardized test situations. Swanson and Howell (1996) also concluded that there is a significant positive relationship between test anxiety and cognitive interference, and a significant negative relationship between test anxiety and study habits. Supplying anxious students with improved test-taking strategies and effective problem-solving skills to replace unproductive worrying seems to work best. Protheroe and Perkins-Gough (2000) suggest to teachers the following methods for reducing test anxiety:

1. Help students prepare ahead of time and provide information about the test. Knowing what to expect increases student confidence.
2. Offer positive reinforcements for student's capabilities and work. Starting the test with a sense of success increases the likelihood of higher test performance.
3. Help students recognize when they are under stress and teach them how to deal with it. Point out that everyone experiences some anxiety during testing, and note that a low level of anxiety in testing can be beneficial to test performance because it makes one alert.
4. If students are taking more than one test in a sitting, give them the opportunity to relax between tests. Having students stretch might help reduce tension.

For younger students, particularly in the early elementary grades, understanding testing terminology, symbols, and procedures becomes critical in comforting them about high-stakes tests. Marzano, Kendall, and Gaddy (1999) found that knowledge of test vocabulary and terminology has a significant impact on student performance on high-stakes tests. Modeling a positive attitude and setting a positive tone for test taking throughout the year will also likely reduce anxiety.

Becoming Assessment Literate

This section describes ways to become assessment literate and how to appropriately use information from high-stakes tests.

Ensuring the quality of assessment data

High-stakes tests are meant to improve student learning and instructional programs, and to enhance public accountability, confidence, and support in the services that school districts offer to students. In order to make the best use of the results, the assessment data must be of high quality. One way to assure quality assessment data is to identify whether the test meets rigorous professional standards. Standards related to the design and use of high-stakes tests are described in *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999). One of the most important standards outlined in this guide is that decisions that will have a major impact on students should not be made on the basis of a single test score, and that other relevant information should be taken into account (see below). It also suggests that multiple test forms be used when there are repeated administrations of an assessment (e.g., a student who fails a test required for graduation should have a reasonable number of opportunities to succeed on equivalent forms of the test); and that the tests used should be technically sound (i.e., reliable and valid).

Utilizing multiple indicators to inform educational decisions

Given their seemingly objective nature (as well as the way they are publicized), there is a temptation for school practitioners to emphasize test scores when making decisions about students or programs. Traditionally, the single indicator used has been a norm-referenced test. Apart from the dangers involved in using results from a single test to make consequential decisions about students or programs, alignment studies in various states show that norm-referenced tests are limited in covering state-adopted standards. Where there is misalignment of this sort, using the test results to make decisions about student learning in relation to the standards could be misleading. With the standards movement, nearly all states are using criterion-referenced tests that are aligned with the state standards ("Quality Counts," 2002). Data from such assessments can lead to better educational decisions because they are based on student performance in relation to the state-adopted standards.

At the same time, the most appropriate approach, particularly when high stakes are involved, is the use of multiple indicators.

Scores from high-stakes tests may prove to be of benefit in decision making if they are coupled with data from district assessments, classroom assessments and observations, and socioeducational factors such as attendance, tardiness, and grade point average. This triangulation of information will help school practitioners make better decisions about students or programs because data from one source can help confirm or disconfirm information from another. The recent revision of eligibility requirements for honors math placement in the Pleasanton Unified School District (PUSD) in California is an illustration of how multiple indicators can be effectively utilized for educational decisions. The PUSD currently uses three different criteria for sixth-grade honors math placement: (a) a student's score on the state-mandated test, (b) a student's score on the district-based end-of-year math test, and (c) a recommendation from the student's fifth-grade teacher (based on student attendance, behavior, effort, and performance on classroom assessments throughout the school year). Using multiple indicators of this sort helps school practitioners cross-validate their judgments about students or programs.

Matching assessment results with the appropriate audience

As shown in Table 3, assessment results are reported at different levels for a variety of purposes. To illustrate, the National Assessment of Education Progress (NAEP) reports data at the state and national level, whereas results for state testing programs may be reported at the student, school, district, and state levels (public reporting of data tends to occur only at the latter two levels). Districts may have their own assessments and these results tend to be reported at the student, school, and/or district level. In addition, teachers utilize numerous types of assessments in the classroom, such as daily assessments, end-of-unit assessments, student projects, and student portfolios. Results for these assessments tend to be reported at the student level.

As shown in Table 3, the level at which results are reported influences the uses to which they

are put. For example, data reported at the student level may be used to provide feedback to students, parents, or both. Classroom-level test scores may be used by a teacher for lesson planning. Becoming assessment literate will increase the chances of identifying the correct audience for which assessment results would be most useful.

Strengths and limitations of different methods for reporting test results

Traditionally, national percentile rank (NPR) scores have been viewed as an easy and convenient way to communicate assessment results to parents, local media, state, and federal officials, and for evaluating the effectiveness of programs in public schools. NPR scores provide valuable information about the performance of individual students in relation to their peers nationwide. For example, if a student's performance on a math test

is described as being at the 70th percentile, this means her score is better than that of 70% of her peers nationwide. Similarly, median national percentile scores are used at the school, district, or even the state level to inform concerned parties about the progress of public schools. Nonetheless, a school's (or student's) increase in NPR scores, say from the 40th percentile to the 50th percentile, does not necessarily mean the same as that school (or student) moving from the 60th to the 70th percentile.

Most norm-referenced standardized tests include in their assessment reports Normal Curve Equivalency (NCE) scores or Scale Scores. These offer a far better estimation of growth than do NPR scores. Whereas the intervals between NPR scores are unequal, NCE scores provide adjustments to the NPR scores so that the intervals are arithmetically the same. The NCE scores range from 1 to 99 with a mean of 50 and a standard deviation of

Table 3
Users, Levels of Reporting, and Uses of Standardized Assessment Results

User	Level of Reporting	Uses
Students	Student Level	Self-evaluation and motivation, focus on content/achievement, and comparison to exemplary work
Parents	Student Level	Home-school communication, timely interventions possible, increased home-school alignment of perceived student achievement, and accurate, understandable, usable information about achievement
Teachers	Classroom Level	Focus on students and learning, targeted reflection, diagnostic planning, individualized student adaptations and/or interventions, and increased accountability data
School-Site Administrators	School Level	Focus on students, teachers, and learning, increased accountability data, targeted staff development, teacher interventions, program and school evaluation and planning
Curriculum Director	District Level	Program evaluation or planning, accountability data, and targeted staff development
Superintendent	District Level	Program or staff interventions, program or system evaluation, and accountability data
School Board	District Level	Administrative interventions, program or system evaluation, accountability data
State Department of Education	State Level	Program or system evaluation, needs assessment and accountability data
Citizen or Legislator	National Level	Accurate, understandable, usable information about schools, district, states, and achievement

Source: Adopted from Jandris (2000)

21.6. Scale Scores, however, range from 1 to 999; they are particularly useful in comparing performance in one subject area across classes, schools, districts, and other large populations; and, especially, in monitoring achievement growth over time.

With the enactment of the No Child Left Behind Act, numerous states are now reviewing their assessments and reporting student performance in relation to state benchmarks and standards in addition to, or instead of, simply reporting norm-referenced scores. Criterion-referenced scores report student performance in relation to a set of designated tasks or skill levels (e.g., the state standards and benchmarks). Results are often presented as performance levels. For example, student performance on the Wisconsin Student Assessment System (WSAS) is reported in terms of four performance levels: Minimal, Basic, Proficient, and Advanced. These classifications describe the extent to which a student is deemed to have met grade-level standards and expectations. Thus, instead of saying that a student's score on the WSAS test is better than that of 70% of his peers, his performance may be described as being at the Basic level in relation to the state standards.

Similarly, the NAEP uses three performance levels—Basic, Proficient, and Advanced—to describe student achievement. In science, the NAEP performance levels are based on a scale from 0 to 300, with 0-138 referring to Basic performance, 139-204 to Proficient performance, and 205-300 to Advanced performance. These performance labels allow for a quick and easy-to-understand overview of student achievement in U.S. schools (e.g., the percent of students scoring at the Proficient level on NAEP in each state). However, one of the limitations of these performance levels (and performance levels in general) is that it is difficult to infer from the results whether most of the students scored at the lower or upper end of the performance level (e.g., a student with a score of 139 and a student with a score of 204 are both deemed Proficient on the NAEP science test). There is also great difficulty involved in establishing cut scores between the performance levels. Because there is no set mathematical procedure for establishing cut points, the process is primarily based on judgment. Many states form panels of judges comprised of

assessment experts, subject area teachers, statisticians, school board members, and citizens to identify the performance levels on the test. Because the choice of cut points is based on the composition of the panel of judges, there is a high degree of subjectivity involved in setting these cut scores. For example, what is termed Proficient in one state may be quite different in another. It may also be quite different from the Proficient level on NAEP.

Understanding norm-referenced and criterion-referenced scores is a step in the right direction toward becoming assessment literate. Using the scores properly will help school practitioners make informed decisions about students or programs.

Error around test scores

In analyzing scores, educators often fail to remember that there is a certain degree of error involved. One reason for this is that a test score is only an estimate of what a student knows and can do, not an exact amount. When test developers design a test, they include only a sample of questions from all possible questions that could be asked in a particular area (e.g., two-digit multiplication). Depending on the questions included on the test, a student may do better or worse. Thus, the score a student receives on a particular test is only one of many she could have potentially received. To illustrate this uncertainty, test developers often provide an estimate of the error in the test. This error measurement can be used to construct a band around an individual's test score. This is done to show that the score a person receives is not the "true" score, but that the true score is likely to be within the band. On tests that contain more error, the band is larger; on tests with less error, the band is narrower.

Why is it important to know about this? Consider, for a moment, a high-stakes test that all students must pass in order to graduate from high school. The test is scored on a scale from 200 to 300 and the passing score is 250. However, the error measurement for the test has been computed and is found to be 5 scale points. If this is used to construct a band around a student's score of 247, the student's true score range would include scores above the passing score. However, this student would be told that they failed the exam on the basis of the score he received.

Other sources of error that are not accounted for by the error measurement described above include extraneous factors such as testing room conditions (e.g., the room is too cold or too hot on the day of the test), student illness (if the student is feeling unwell, they may not do their best), or flaws in the test (a poorly written question with no correct answer, or with more than one correct answer). Taken together, these various sources of error suggest that careful interpretation and use should be made of test results, particularly when they are used for high-stakes decisions.

Conclusion

School practitioners need to adequately and appropriately prepare students for high-stakes testing without detracting from real learning. Teaching to the state standards and using various assessment approaches and formats constitutes appropriate test preparation practice. When educators understand test preparation practices from an instructional perspective, the integration of appropriate test preparation practices into regular classroom instruction becomes fairly easy. This allows school practitioners to better focus on student learning rather than just raising test scores.

Education leaders need to be assessment literate in order to respond to the demands of the avalanche of high-stakes testing. Being assessment literate broadens one's perspective to view assessment as a dynamic process. The effective use of assessment results includes utilizing multiple modes of assessment and examining longitudinal data for educational decisions. The process involves taking picture-perfect snapshots over time and constantly compiling a photo album about student learning or program effectiveness. High-stakes tests should be viewed as tools to help take some of those snapshots. The goal is to gather a variety of information to best inform educational decisions rather than limit our judgments to a single assessment. Educators must become knowledgeable about appropriate test preparation practices and prepare their students for this new world of high-stakes testing.

References

American Educational Research Association, American Psychological Association, & National Council on

- Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Berliner, D., & Casanova, U. (1988). How do we balance test anxiety and achievement? *Instructor*, 97(8), 14-15.
- Clovis, D.L. (1999). Take out your no. 2 pencils: Taking the stress out of standardized tests. *Scholastic Instructor*, 108(7), 27-28.
- Crocker, L., Schmitt, A., & Tang, L. (1988). Test anxiety and standardized achievement test performance in the middle school years. *Measurement and Evaluation in Counseling and Development*, 20(4), 149-157.
- Chittooran, M.M., & Miles, D.D. (2001, April). *Test-taking skills for multiple-choice formats: Implications for school psychologists*. Paper presented at the annual meeting of the National Association of School Psychologists, Washington, DC.
- Guthrie, J.T., & Wigfield, A. (2000). Engagement and motivation in reading. In M.K. Kamil, P.T. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Volume III* (pp. 403-422). Mahwah, NJ: Erlbaum.
- Hancock, D.R. (2001). Effects of test anxiety and evaluative threat on students' achievement and motivation. *The Journal of Educational Research*, 94(5), 284-290.
- Haydel, A.M., & Roeser, R.W. (2002). *On the links between students' motivational patterns and their perceptions of, beliefs about, and performance on different types of science assessments: A multidimensional approach to achievement validation* (CSE Technical Report No. 573). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Jakupcak, J., & Rushton, R. (1992). *Corvallis school inclusion project* (Report No. 141). (ERIC Document Reproduction Service No. ED357566).
- Jandris, T.P. (2000). *Essentials for principals: Data-based decision-making*. Alexandria, VA: National Association of Elementary School Principals.
- Loulou, D. (1997). *How to study for and take college tests* (Report No. NLE-97-2527). (ERIC Document Reproduction Service No. ED404378).
- Marzano, R.J., Kendall, J.S., & Gaddy, B.B. (1999). *Essential knowledge: The debate over what American students should know*. Aurora, CO: McREL Institute.
- Mehrens, W.A. (1991, April). *Defensible/indefensible instructional preparation for high stakes achievement tests: An exploratory dialogue*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Miyasaka, J.R. (2000, April). *A framework for evaluating the validity of test preparation practices*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

- Norton, S.M., & Park, H.S. (1996, November). *Relationships between test preparation and academic performance on a statewide high school exit examination*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Tuscaloosa, AL.
- Protheroe, N., & Perkins-Gough, D. (2000). *Essentials for principals: Meeting the challenges of high-stakes testing*. Alexandria, VA: National Association of Elementary School Principals.
- Quality counts 2002: Building blocks for success [Special Report]. (2002, January 7). Bethesda, MD: *Education Week*.
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23(3), 197-227.
- Roth, J.L., & Paris, S.G. (1991, April). *Motivational differences in students' perceptions of classroom and standardized achievement tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Smith, R.J., Arnkoff, D., & Wright, T. (1990). Test anxiety and academic competence: A comparison of alternative models. *Journal of Counseling Psychology*, 37(3), 313-321.
- Swanson, S., & Howell, C. (1996). Test anxiety in adolescents with learning disabilities and behavior disorders. *Exceptional Children*, 62(5), 389-397.

TIP