Appendix A Continued

(3) Preparation practices of other teachers
"Teachers design their instruction to prepare students for standardized tests."

(4) Parents’ misuse of standardized test results
"Parents often use standardized test scores to determine the quality of the school."

(5) Frequency with which teachers give their students help to prepare for standardized tests
"How often do teachers in your school give general tips on how to take tests?"

REFERENCES

The Immorality of Test Security

GRANT WIGGINS

In this essay the author argues that test security in student testing is a practice of questionable morality and efficacy, despite the long-standing use of the practice by test-makers. Arguing that the practice is an unhinging immoral habit, with its roots in premodern medieval views of the teacher-student relationship, the author proposes that new policies and practices be formulated to protect the inherently vulnerable student in the same way that laws now protect adult test-takers and experimentees. The author argues that secrecy prior to, during, and after a test be scrutinized carefully and minimized through the use of explicit principles—an assessment Bill of Rights—that put the students' rights on a par with those of the test-maker and that honor the modern rights of informed consent and due process.

It is so common that we barely give it a second thought. The tests designed to evaluate the success of student learning invariably depend on secrecy: secrecy with respect to the questions that will be asked; secrecy about how the questions will be chosen; secrecy about how the results will be scored; secrecy as to what the scores mean (as in cases when we are not given back our tests and an answer key); secrecy as to how the results can be used. What a paradoxical—and ultimately immoral—affair! Our aim is to educate, prepare, and enlighten; our habit of testing is built on procedures that continually keep students in the dark.

The legacy of secrecy in student testing is long-standing. It is rooted in the world of autocratic religious power and the hierarchical guild mentality of the Middle Ages—a world filled with many adult secret societies (Bok, Educational Policy, Vol 8 No 2, June 1994 157-182
© 1994 Carvin Press, Inc.

157
1983). The secret test as we know it came from this world, a culture where professors saw themselves as members of a guild; a world where status transcended rights and was granted by degrees to only a few; where the assessor had the unquestionable right to demand (orthodox) answers of the test taker without justifying either the questions asked or the evaluation of the answers. Whatever modern justifications are made for test security on validity, proprietary, or efficiency grounds, there is little doubt that the roots of the practice derive from this premodern, assumed right to keep vital information from the examinee.

In this essay, I call attention to the questionable premises on which secrecy in student testing—test security—is based, and suggest principles for better distinguishing between appropriate and inappropriate use of secrecy in assessment. Those who rely on secrecy are presumed to require justification: not knowing how one will be assessed and not having the ability to question test results is prima facie contrary to due process. What policies should thus protect the student from inappropriate practice used on technical or efficient grounds only? A great need exists at the policy level for an intellectual Bill of Rights: principles that would protect student test takers (and test users—teachers—in the case of mandated large-scale school testing) from dubious practices. Time-honored forms of secrecy prior to, during, and after an academic test serve neither test takers and test users (teachers) nor policymakers (who view tests as "accountability") if the aim of testing is not merely to audit performance but to improve it (Wiggins, 1993).

The Moral Issues Raised By Test Security

Careful examination of technical guides to enforcing test security alert one to the danger of unthinking secrecy in student assessment. For example, in an old textbook on testing we are advised: "Deny the examinee information concerning the rightness or wrongness of his response" as a way of keeping the test secure for other possible test takers (Highland, 1955). Yet even less obviously objectionable practices, such as the constant use of indirect items and arcane scoring processes, render testing mysterious and have a moral price. Such secrecy keeps most test takers from understanding how they did what might be done about it, and effectively limits any critical complaints from either student or teacher concerning items and adverse impact. An unrelenting dose of proxy tests also keep test takers in the dark about real performance tasks and standards—an equity issue with great consequences because in poor schools students never come to grasp what kinds of performances the adult world values beyond school walls (Wiggins, 1993).

The illusion of objectivity in indirect testing obscures matters further. The items and their point value are of course proposed by human judges even if the scoring is mechanical; such judgments are obscured when tests are secure. Fairness and equal oversight are clearly threatened when any judge does his or her work in secret. Imagine the harm to our political system and citizen insight into, and faith in, our principles if Supreme Court judges did not have to publish opinions to support their decisions. Consider how much better tests might be if every test and answer key had to be justified in writing in the same way that court decisions are made.

Language obscures moral matters as well. We need to be reminded that the phrase secure test is a euphemism. Security implies that the test maker has the only key moral claim: a proprietary interest in keeping the test secret; a priority that outweighs all other claims in a society organized on Lockean principles where property rights are fundamental. But the student's (and teacher's) more basic right to question is at stake in all testing relationships. Secret tests make it far more likely that the relationship is morally imbalanced. This is clearer when we consider what would likely happen if some company invented a secure performance appraisal system for its employees. It would no doubt be rejected as immoral and illegal.

We therefore need to ask some elemental questions to remind us of what is at stake. Why do we take for granted that student test takers do not have a basic right to prior knowledge and justification of the form, content, and standards by which their work will be judged? Why would we presume that a test designer's proprietary rights always outweigh the test taker's moral right to due process? Why would we assume, contrary to all accepted guidelines of experimental research, that educational testers need not publish their tests and results after the fact for scrutiny by independent experts as well as the test taker? As with all blind spots, we fail to see how our test rituals sanctify a stance that is at best morally questionable, at worst self-serving but rationalized by custom—like racial discrimination.

We have a duty to ask, then: should the assessor be assumed to have the right to employ such secrecy, irrespective of its impact on students, teachers, and schools? Or is the tester always obligated to justify the practice and its extent, and to minimize its use? I argue the latter, yet I am not sanguine about the near-term possibilities. This pessimism derives from the fact that we are dealing with a habit, not a mindful policy; and we are dealing with a clash between the rights of test-making adults and test-taking children. All the more reason to consider the need for conscious and explicit policies having to do with test secrecy.

As with all deep-seated habits, we do not see the irrationality of our reliance on secretiveness in educational testing beneath our many rationalizations. We say, "Surely we want the student to be able to grapple with the unexpected.... We couldn't possibly rate each candidate efficiently and
effectively any other way. . . How could we *possibly* give the test taker access to the questions in advance without corrupting the test? . . . There is no way to obtain validity without test security. . . . Hasn’t testing *always* been done this way? Isn’t our testing system more ‘fair’ since everyone is judged in the same way?” The unthinking character of these types of responses would be more plain if it weren’t for the ubiquity and seductive ease of use of secrecy.

**Different Kinds of Test Secrecy Analyzed**

Given the thoughtless reliance on secrecy in testing, and the obvious harm such secrecy poses for rehearsal, performance mastery, and later improvement, it is worthwhile to think through what we mean by calling a test secret. In what sense is a test a secret and in what sense must a test be secret for it not to be compromised? Whether or not a test is compromised by being no longer secure, when is such secrecy unfair to the test-taker or test-user? Where do the test-taker’s rights become predominant, in other words, in a conflict of interest regarding security in testing? A variety of practices are considered in which the testing that takes place in schools is predicated on a degree of secrecy about what will be assessed, what methods will be used, the value of those methods, what the results mean, and/or the value of the test results. This article is limited to testing in academic situations, because in school testing the moral claim for openness is clearly greatest: it must be educative, not just precise or efficient.

Everyone has experienced the most obvious answer to the question of what a secret is in conventional student testing. Each of us has encountered many tests where we entered the room not knowing the specific questions that were going to be asked. Such a practice is so common that it has a profound hold on our consciousness. “Secret up until the day it occurs”—that’s what a school test is. The mythic power of this practice was made very clear in a story told to me by Sharon P., a student teacher who tried some novel approaches to instruction and assessment in her student teaching practicum. She designed a culminating unit in her English course where the students, in small groups, devised all the would-be questions to appear on the final exam, according to certain criteria supplied by Sharon. The reaction of two teachers in the faculty room at her school summed up the irrational hold the practice of test security has on us: “But then the students will cheat!” they protested in unison. Nor were they convinced of the wisdom of such a policy even after a lengthy discussion.

Prior secrecy is not the only constraint in testing openness, however. There are 11 kinds of secrecy built into conventional testing that occur before, during, and after a test. Of the prior-to-testing kind, there are three:

- Knowledge of what *specific* questions will be on the test is usually a “secret” known only to the test maker.
- Knowledge of the possible questions that might be on the test (i.e., the student may or may not have access to the set of possible questions or to the database described above, a larger known set from which the later secure test will sample.)
- The time and timing of the upcoming test can sometimes be a secret, a common feature of the high school or college quiz—usually described euphemistically by the teacher on the day in question as announced.

The following kinds of mystery, hiddenness, or constraint on knowledge might be thought of as secrecy during the test:

- The test itself can be secret in that the student is actually unaware that a test is taking place. (In the technical literature, this is called “unobtrusive” testing.)
- The scoring criteria and standards can be secret (when the questions are not multiple choice or equally weighted in value) in that the student is unaware of the scoring criteria, descriptors for the range of answers, and the anchor performances being used to set standards.
- Whether one is on the right track or not in conceiving of a question or an approach to answering it is a secret. The student cannot typically converse with the test maker or test administrator; in other words, cannot check on the meaning of a question or the aptness of an approach to answering the question.
- The resources for constructing and double-checking possible answers are kept a secret (to stretch the term a bit) in the sense that the real-world resources that would be appropriately available are kept explicitly hidden from the student during the test. This type of security extends to the use of other people as resources, including not only students but any adults present.

The following kinds of secrecy refer to what is not revealed after the test:

- The meaning of the results is kept secret when the tester does not release the actual test paper from which one can confirm or critique one’s score, and the test itself. As mentioned in the previous chapter, for example, it is still common in many undergraduate exam situations for students to not receive back their “blue books” after the exam.
• The test can remain secret in that even if the students get their papers back, they may not get to see the answer key or sets of exemplary test papers, against which one is to compare their answers.

• The significance of a national or state multiple-choice test is a secret by virtue of its indirect nature. It almost never has obvious relevance to local curriculum or face validity to the student and teacher. And when technical validation for the test tasks and standards does exist, it is rarely meaningful to the student and the teacher.

• The value of the test for future learning and aspirations remains a “secret” in that most testing is one-shot, not scaled using longitudinal scoring criteria and standards, and composed of indirect items.

What we seek are some principles for judging the boundary lines. In what contexts is each secrecy appropriate and in what contexts is it not? How might we better judge when any particular use of secrecy in assessment goes over the line and becomes unjustified? There may well be times, in the classroom as well as at the state and national level, when test security can be justified. What can no longer be justified, however, is the unspoken view that psychometricians and teachers can assume a right to make such secrecy a cornerstone of test methodology.

The technical argument for secrecy about what will be tested is fairly straightforward. The validity of all short-answer tests (though not necessarily examinations or most authentic assessments) is typically compromised if the test questions are known in advance. Let us recall why this is so. Many questions, if known in advance, would enable the student to short-circuit the learning process. The student would now be in a position to know the correct answer without necessarily knowing why it is so, given the format. I can, for example, memorize the questions and correct answers, perhaps even gaining the right answers from someone else. But this result would render invalid the inference that the student who does well on the test “knows and understands” the subject matter tested.

Second, insight into the student’s breadth of knowledge is perhaps compromised by undoing secrecy about the questions before the test. Because most tests involve a necessarily small sample of the total domain of a subject, and what was taught, knowing the questions in advance can make the student wrongly appear to have a very expansive knowledge. Put differently, in most testing situations we want and expect to be able to say more about the breadth of mastery than just what was tested, as when we construct samples of 1100 for Gallup polls to be generalizable to the whole nation. Instead of making an effort to successfully gain control over the domain of all possible important tasks or questions, the student need only concentrate on the small sample of questions now known to be on the test. Of note, therefore, is that a test of a few essay questions or performance tasks may put the teacher or test maker more at risk for such an invalid inference, if the essays are the entire test. Since the student need only prepare for one or two essay questions—versus the typical hundred or more on a multiple-choice test—the validity of any inference about student control of all the important topics may be severely compromised by advance knowledge of the questions.

The Emperor, Revisited

The seeming necessity and the commonness of the practice of keeping student test takers in the dark about tests is so much a part of the educational landscape that a vivid story may be required to cast the problem in a fresh, revealing light. We might do well, therefore, to consider our unthinking use of secrecy in testing as a modern version of the Emperor’s New Clothes (Lewis, 1981).

You no doubt recall the story: rascals pose as tailors and “weave” a suit of the finest cloth for the king, earning riches by the fashioning of an illusion. The king’s nakedness, there to be seen by all, remains “unseen.” A sham that should be obvious works precisely because of the tailors’ warning: only those who are stupid or unfit for their positions would be unable to see the cloth. And so it happens that the townspeople rationalize their perceptions of nakedness and their secret doubts; they, like the king’s retinue who fear for their honor, praise the king as he parades in his “finery.” The king, too, although doubting that he is stupid or unfit, is sucked into the self-deception. It is the innocent child, unaware of the “secret” or the need to maintain secrecy, who exposes the hoax. “But he has nothing on!” exclaims the child.

Few usually recall the story’s ending. The townspeople don’t immediately come to their senses; the elders initially dismiss the remark of the young innocent. Eventually the child’s words are repeated enough so that their truth cuts through the self-deception and doubt. But, while thinking that the now-skeptical townspeople must be right, “the Emperor thought to himself, ‘I must not stop or it will spoil the procession.’ So he marched on even more proudly than before, and the courtiers continued to carry a train that was not there at all.”

The tale is instructive about current testing policy on many levels. We are still dismissing the remarks of the “innocents.” We do not look through the eyes of student test takers as they prepare for and take the tests we use to see how debilitating they are to intellectual engagement, courage, and imagination. Nor do we look over the shoulder of employers, teachers and administrators to see how rarely they use test results or understand the meaning of
test-taker scores and errors. Commercial test makers literally profit from the illusion that, like the tailor’s yarn, all “fine” tests must be built with a specialist’s mysterious skill. Testing, rather than being a very common and above-board practice of assessing performance on known tasks we value, against known and apt criteria, becomes an arcane science that is entrusted—and apparently only entrustable—to statisticians. Critics of such tests fear looking like the crude folks the tailors warn their critics will be; wary practitioners are routinely made to feel ignorant of the true “finery” in test validity and reliability.

The story also shows how a secret attached to power can cause self-deception and the weakening of faith in judgment. The hiddenness of test items, and their utter simplicity when they are seen in the test setting, is much like that of the king’s nakedness. It is so obvious yet so mysteriously generated; and encountered in enforced silence, so as to make one feel certain that arcane standards “must” render the seemingly nonexistent test “garment” substantive. Like the townspeople in the story, everyone from teachers to superintendents to test takers can end up talking as if the real challenges (or “criterion performances,” in psychometric talk) we value were being used to directly observe academic performance and achievement. The supposed necessity of keeping test questions secure eventually becomes obvious to everyone—as if we have forgotten that almost all the important tests and criteria in life (for getting employed and getting a raise, writing a successful dissertation for a doctorate, submitting a winning engineering design bid, successfully trying a case, curing a patient, or passing legislation) also involved secret tasks, criteria, and standards. The mystery of test maker authority ensures that private doubts remain inchoate; highfalutin’ (but hazily understood) technical language comes to dominate our conversation.⁴

Although teachers talk of the foolishness or harm of secure indirect tests, they end up employing their own versions of them—the true sign of the tests’ mythic rather than rational, power. The esoteric (but misunderstood) procedures of the tailors take root and are poorly mimicked by the uninhibited: any inspection of local tests reveals that almost all of them are of questionable validity and reliability (Stiggins, 1991). The call for more sound standardized tests then naturally increases from the outside critics; the vicious circle continues.⁵ Bring in the tailors! Let the king march more proudly! But pity the poor student. For, unlike in the story, the child’s voice—common sense—remains unheard or dismissed still.

The One-Sidedness of Secrecy

As Sissela Bok (1983) stressed in her fine work on the subject a decade ago, a systematic use of secrecy “can debilitating judgment” because it makes it possible to “shut out criticism and feedback, leading people to become mired down in stereotyped, unexamined and often erroneous beliefs and ways of thinking.” In professional, personal and social matters, the principle of mutual respect must not be at risk in any use of secrecy. To be kept in the dark is to be treated disrespectfully—as an object, in moral language. That is after all why informed consent is now the moral cornerstone of any attempt to experiment with human subjects. The importance in being explicit about such respect should be clearer in light of the recent revelations about experiments involving radiation on unwitting patients at the dawn of the atomic age. Whether it be patriotism or education, someone in power will invariably claim that someone’s keeping of secrets is in someone else’s interest. But the burden of proof in a modern democracy is on the secret-keeper, given that the keeper of secrets has unilateral power to decide for us what is in our interest. “This is not to say that some people might not be granted limited powers for certain purposes . . . but they would have to advance reasons sufficient to overcome the initial presumption favoring equality” (Bok, 1983).

This obligation to be more open with students about tests (hence, open to their scrutiny, complaints and needs) might be easier to see if we think of secrecy in relation to citizens. The right to scrutinize control over what is kept secret becomes more compelling at the level of institutions, where power is even greater and abuse is more likely: “When power is joined to secrecy, the danger of spread and abuse . . . increases. In all such cases the presumption shifts [away from the assumption of a right to secrecy]. When those who exercise power . . . claim control over secrecy and openness, it is up to them to show why giving them such control is necessary and what kinds of safeguards they propose. . . . Even where persuasive reasons for collective practices of secrecy can be stated, accountability is indispensable” (Bok, 1983).

Our unwillingness to tolerate such one-sided secrecy when more obvious (adult) rights are at stake becomes clear when we examine the policies that have arisen when employees are being tested using secure tests for hiring or promotion. Irrespective of the wishes or habitual practices of testing specialists, the courts have been quite clear that the assessee has a right to more information (and more stringent, user-friendly tests of validity) than testers have historically wanted to provide (Berk, 1986).

Such a principle has always been difficult to honor when dealing with children or others who seem morally inferior to us, however. Yet progress has been made. We now take this right for granted in judicial inquiries. Children can be legally protected from abusive or neglectful parents; they can sue for “divorce.” More generally, openness in terms of evidence and the ability to
cross-examine in all formal disputes is a cornerstone of the modern judiciary and requiring that secrecy or one-sided knowledge always be minimized, even with respect to evidence. But it was not always so. The insistence on mutual respect and openness in formal inquiries is recent, as Foucault (1977) notes. His history of legal investigation and examination in 17th-century France is a reminder of why moral equality is so difficult to uphold in practice.

In France, as in most European countries, with the notable exception of England, the entire criminal procedure, right up to the sentence, remained secret: that is to say, opaque, not only to the public but the accused himself... Knowledge was the absolute privilege of the prosecution. The preliminary investigation was carried out as “diligently and secretly as may be,” as the edict of 1498 put it. According to the ordinance of 1670, which confirmed... and reinforced the severity of the preceding period, it was impossible for the accused to have access to the documents of the case... impossible to know the nature of the evidence... impossible to make use of the documents in the proof. (Foucault, 1977, p. 35)

The combination of secrecy and unilateral respect could even justify deceit in the judge’s conduct that we would now find morally repugnant:

The magistrate, for his part, had the right to accept anonymous denunciations, to conceal from the accused the nature of the action, to question him with a view to catching him out, to use insinuations... The secret and written form of the procedure reflects the principle that... establishment of truth was the absolute right and the exclusive power of the sovereign and his judges. (Foucault, 1977, p. 35)

It is not pushing the argument too much to ask the reader to re-read the passages and think of the student as the “accused” and the test-maker as the judge here. (Knowledge as the privilege of the assessor... impossible to know the nature of the evidence... to question him with a view to catching him out—that is, the use of distracters in tests, and so forth.) At the heart of Foucault’s analysis is the view that such one-sided practices have been common to all the areas where we seek to “discipline” humankind (law, military, education, and psychiatry). Foucault (1977) explicitly links the judicial and educational “examination” that “combines the technique of observing hierarchy and those of a normalizing judgment.” Although he argues that “investigation” has become modernized through the methods of science, the “examination is still caught up in disciplinary technology.” The judicial metaphor is apt here in another sense: One might think of students as “guilty” of ignorance until proven “innocent” by evidence from tests.

To link these observations to modern school testing practices, Sue Barton (1990) notes, for example, that tests are “obscure, opaque, inaccessible in many of the same senses. She also observes that “these tests and their results carry no intuitive meaning to anyone besides educators... They thus fail to measure objectives that parties with interests in the outcomes of our educational system can understand, ‘see,’ and debate.” To try to talk with a parent about stanines and grade-level equivalents or standard deviations is to experience this problem first hand: They do not grasp what scores mean, and they tend to assume that scores are infallible. Secrets (and one-sided power) are maintained not merely by overt furtiveness but by a constant use of esoteric language that only a few can penetrate.

The Dysfunctional Nature of Secrecy in Educational Testing: Some Revealing Vignettes

Secrecy is not merely morally problematic. In educational testing it is downright counterproductive, as the fanciful scenarios that follow make clearer.

Vignette 1. Imagine the students to be the managers of a warehouse, where organization of incoming material and ability to pull together future orders are paramount. But reflect on what the manager’s job might be like if he or she knew neither what kind of material would be arriving each day, nor the quantity, nor the kinds of orders that would later have to be filled. This is the student’s position in school.

Each day, new material arrives, sometimes too quickly for the student manager to organize it on shelves. After a week of ordering the contents, new and unexpected material arrives that compels the student to completely rethink the system used in storing—with no assurance that the revised system will compensate for future, unknown deliveries.

After months of storing and organizing catch-as-catch-can, the student managers are warned by the central office that they must be prepared to correctly fill dozens of (unknown) orders on the spot, without access to the notes and resources that serve as their database. The managers will be “tested” on their ability to predict the order, and fill it from memory—not the more authentic and vital ability to plan for, refine and routinely process a wide range of known and varying orders.

A simpler image of the foolishness of this approach, if our aim is mastery, shifts the focus to music performance testing. Imagine having to play a musical piece never seen until the last day of music class, in which you cannot hear how you are doing as you are tested, and where a score comes back a few weeks later. It is hard to imagine how the claims for so-called accountability provided by one-shot secure tests can even be justified when we consider how poor the feedback loop is in such a system: No one can learn
or adjust from the results of such tests; yet year-end, secure and large-scale state tests are supposed to provide accountability using such methods.

**Vignette 2.** Consider how the combination of security and the use of proxy (indirect) secret items ultimately corrupts performance. What would happen if baseball were played all season long, but the pennant races were decided using one-shot, secure tests with an aggregate score designed by statisticians? Thus, on the last day of the season, specially constructed—and secure—tests would be given to each player, composed of static drills and game situations. The pennant races would be decided by each year's (new) test and its results. Who believes that this secrecy, so necessary to the test's validity, would not end up corrupting both the game and the players' motivation? (Note that the students' current situation is actually worse, because they are usually not allowed to "play the game" of authentic knowledge use and production, but must endure syllabi composed of drills and game situations, ordered in "scope and sequence" fashion. Not even learning the "game" of knowledge in use, the students are even less likely to predict this kind of testing situations they will face.)

More than the first vignette, this one reveals how unwittingly and inappropriately powerful the possessors of secret testing knowledge and criteria can be—even if their aim in life is to be merely helpful statisticians. The test designer here supposedly seeks only to design a valid sampling test of all the important elements of performance as specified by others. The secrecy is justified because without it the test is corrupted: coaches would "teach the test"—in the bad sense of short-changing the players' overall education so as to artificially raise the score and distort its meaning. Yet, we easily see how such a system would corrupt coaching and the game itself anyway. Not only the student-players, but the teacher-coaches would be robbed of the capacity to concentrate on the time-consuming task of developing excellent play beyond drill in discrete skill—just as in the classroom teachers never end up asking students to use knowledge in complex, integrated ways, given the demands of preparing for the multiple-choice test (Fredriksen & Collins, 1989). This vignette accurately reflects how the SATs and other high-stakes tests have corrupted teaching and learning.

**Confronting the Arguments for Test Security Before, During, and After the Test**

As mentioned above, the traditional defense for prior secrecy about what will be tested has to do with validity: If the student knows the questions in advance, the inferences we wish to make about depth and breadth of knowledge are rendered invalid. This is a solvable problem, however, if the student is given access to a representative set of possible test questions, and if we acknowledge that most tests overemize on low-level information-focused questions. The model of having a set of possible questions in advance is, after all, a common feature of graduate and undergraduate education. One studies all the questions carefully—a methodical and comprehensive review of the priority questions of the course—and knows that some of those questions will comprise all or most of the final examination. Some questions would admittedly go unasked if this were to be a common practice, but we would do well to consider why we feel the need to ask those kinds of questions.

There are certainly shades of gray here. If we say that the questions should be secure, do we hold the same view about the type or format of questions to be asked? At present it is considered sound policy for teachers or testing companies and agencies to describe the types of questions to be asked (definition/true-false/essay, for example), and to provide samples of each type of question for review. It is just this kind of modest "de-classifying" that enables SAT and ACT preparation services to function and thrive. If this is deemed a fair practice, then why wouldn't we offer some opportunities for students to practice some of the particular questions, for example, as a take-home part of an exam? It would seem unfair on the face of it for the student not to know the type of question and its approximate importance in the whole test; that would appear to violate basic principles about due process and equal success. (And a failure to ensure prior access to the test format might upset the technical validity of the results, if some are familiar with the format and others are not.) Similarly, it would seem inappropriate for the student not to know the weight of a subscore, or the criteria, or standards being used when the test involves judgment-based scoring. How would I know how to apportion time or judge the length or depth of my response without such information?

Although post facto test disclosure has been debated for years, only recently have arguments and feasible proposals been made for greater openness prior to a test being taken, including large-scale tests. Judith Schwartz of Harvard has been among the most vocal and eloquent in arguing that the complete item bank from which any test might be constructed also be made public, following up on an idea made by J. Zacharias in 1979. Given that any test is a sample; and given that the item bank can be made large enough to encompass the domain of possible questions and thus prevent narrow cramming and invalid inferences about results, there is little reason to maintain perpetual security, and many good reasons for making that bank of questions open for public inspection. We could easily develop "large publicly available data bases of reviewed and signed problems in a wide variety of school subject areas. . . . School systems and state boards of education and other
responsible authorities would use these data bases as the sources of the questions and problems in the accountability [systems]” (Schwartz & Viator, 1990).

In the example provided, Schwartz proposes that the state or district make clear which sections of the data base would be tapped, and what the criteria would be for the selection of particular questions. The database would be so available that it could be found in all “libraries and bookstores” with the clear advantage of such a “sunshine” procedure:

If the pool of problems from which examinations are composed is publicly available in advance, then any group in society that feels ill-served by the substance or language or context of a question can raise a prima facie objection in advance of the use of the question and not have to rely on the vagaries of statistical analyses to claim bias or prejudice after the fact.

A regional, state, or national database would enable us to have far more open testing, without compromising validity, if we define openness as prior knowledge of the complete domain of all specific questions or tasks that we value and have in our collective possession. Then, it is no longer feasible to cheat on the validity issue because it is not practical to master each question in advance; one would still have to worry about the entire domain of knowledge that the test is tapping. In many cases at the high school and college level all possible test questions are actually known in advance by course; what is “secret” is which few from the larger list will actually be used.

Indeed, if one thought of the portfolio as a set of products and performances meeting certain criteria with respect to range of genre, topic, or type of product, then it would be quite possible to dispense with most security. Consider the following examples of performance-based course requirements:

- Interesting-to-students history of the sites
- Annotated bibliography: recommended readings for other students

b. The writing of an International Bill of Rights
- Refer to past attempts and their strengths and weaknesses: Helsinki accords, Communist Manifesto, U.S. Bill of Rights, etc.
- Student must obtain the “signing on” of a diverse group of peers and adults.

c. Report to the secretary of state: Policy analysis and background report on a Latin American country: what should our short-term policy goals with the country? What are its economic and political prospects?
- Collecting and analyzing media reports from other countries on U.S. policies in the Middle East
- Put together a “briefing book” of photocopied press clips for the president, with commentary on accuracy of story
- Videotape/audio-tape a simulated newscast summarizing world reaction to a recent U.S. policy decision

d. Conduct an oral history on a topical but historically interesting issue:
- Recent American immigrants
- Veterans of “Desert Storm,” Vietnam, and World War II on American role as “policeman” in world affairs

- Design a museum exhibit, using artifacts and facsimiles:
  - On links between a European country’s geography and its economy
  - On local role in the Industrial Revolution
  - On patterns of modern emigration and their causes

e. Write and deliver, on videotape, 2 speeches: by the visiting head of an African country on the history of United States-Africa relations, and a response by President Clinton’s spokesman.

f. Take part in a formal debate on a controversial issue of global significance, e.g., aid to Russian republics, the U.S. role in the fall of communism, etc.
- A model UN: groups of 2-3, representing a country: a new Security Council resolution on terrorism
- Write a textbook chapter: “People, Ideas, Events, or Economic Conditions: Which are the primary causes of revolution? Were the most important revolutions ‘revolutionary’ or ‘evolutionary?’”

Major Tasks for a Global Studies Course

a. The design of a tour of the world’s most holy sites
   - Accurate maps included
   - Guidebook with description of local norms, customs, etiquette
   - Analysis of most cost-effective route, means of transportation

b. The writing of an International Bill of Rights
   - Refer to past attempts and their strengths and weaknesses: Helsinki accords, Communist Manifesto, U.S. Bill of Rights, etc.
   - Student must obtain the “signing on” of a diverse group of peers and adults.

c. Report to the secretary of state: Policy analysis and background report on a Latin American country: what should our short-term policy goals with the country? What are its economic and political prospects?
   - Collecting and analyzing media reports from other countries on U.S. policies in the Middle East
   - Put together a “briefing book” of photocopied press clips for the president, with commentary on accuracy of story
   - Videotape/audio-tape a simulated newscast summarizing world reaction to a recent U.S. policy decision

d. Conduct an oral history on a topical but historically interesting issue:
   - Recent American immigrants
   - Veterans of “Desert Storm,” Vietnam, and World War II on American role as “policeman” in world affairs

- Design a museum exhibit, using artifacts and facsimiles:
  - On links between a European country’s geography and its economy
  - On local role in the Industrial Revolution
  - On patterns of modern emigration and their causes

e. Write and deliver, on videotape, 2 speeches: by the visiting head of an African country on the history of United States-Africa relations, and a response by President Clinton’s spokesman.

f. Take part in a formal debate on a controversial issue of global significance, e.g., aid to Russian republics, the U.S. role in the fall of communism, etc.
- A model UN: groups of 2-3, representing a country: a new Security Council resolution on terrorism
- Write a textbook chapter: “People, Ideas, Events, or Economic Conditions: Which are the primary causes of revolution? Were the most important revolutions ‘revolutionary’ or ‘evolutionary?’”
One uncontroversial implication of this kind of openness concerning the domain and the sample is that we would no doubt improve the quality of all tests and test questions. If test-makers, including teachers, had to publicly provide a (perhaps large) pool from which the test would sample, the current preponderance of low-level recall questions having little to do with genuine competence in a field of study would be undercut. Thoughtful and deep understanding is simply not assessable through only multiple-choice testing. We will continue to send the unwitting message to teachers that simplistic recall or application, based on coverage is all that matters until we change this policy.

In short, we should not let the test-maker off the hook on the most basic form of openness: prior knowledge of the questions themselves. Knowing the specific test questions or tasks in advance is the norm in professional life. Similarly for student actors, musicians, athletes, and science fair goers. It might engender more chaos and greater cost in testing as we know it if tests were not secure. But we should at least explore a cost-benefit analysis of it, given how clearly it is in the student's interest for tests to be as open as possible.

Declaring that any educational test should be as open as possible does not take us very far in determining the limits but at least makes the matter more explicit, subject to scrutiny and external review. Such a policy would compel test designers to think through the matter of the student's right to proper rehearsal more responsibly. It would also likely lead to tests that have different parts, some of which are appropriately known in advance and others that are not. (We see such a model used all the time in performance competitions such as Odyssey of the Mind, formal debates, music competitions, and so forth.)

It is, of course, the case that in fields involving the vagaries of human interaction, one cannot know precisely what the test will require: there are inherently unanticipatable aspects to real-world tests. The client does the unexpected; the other team seeks to outwit us; the pilot does not know the precise maneuvers that will be required in difficult conditions, and so forth. Performance testing surely demands that we test such adaptiveness (a capacity that Binet viewed as the hallmark of intelligence). Some would use this criterion as an excuse for retaining complete prior test security, but the real-world situation is quite different: the adult always has the opportunity to adjust, consult resources, ask questions—even bring in others or seek a delay in order to consult—while the test is in progress. The diligent student can rehearse the task effectively even if in the real test the facts, problems, or contexts are varied or unpredictable as long as the likely moves are taught and tested. This is of course what coaches of athletes, performing artists, and teachers of medicine and law help students do: know the difference between what is inherently anticipatable and what is not, work on what is, and train for the imaginable unanticipatable events. As General Schwartzkopf, the Commander of "Desert Storm," put it in describing real tests: assume that the best will happen, but be prepared for the worst.

Secrecy During the Test

The rationale for standardizing test conditions and materials is clear enough: Fairness and validity require that each student have equal opportunity to answer a question correctly. If I do not standardize what resources can or cannot be used; if I do not develop a standard protocol for what test administrators can or cannot say, then we run the risk of invalid inferences about performance and unreliable scores.

But that rationale hardly justifies the regular practice of forbidding almost all human interaction and the use of contextually appropriate resources, particularly if our aim is to make tests educative and more authentic, and to determine whether students can intelligently use resources.

With respect to the secrecy of resources during the test, the technical constraints requiring the standardizing of test conditions invariably seem to inappropriately outweigh the pedagogical ones. Keeping the essential tools of research and reference unavailable compromises the construct validity of a test. If the nature of the simplified items used in tests prevents the use of books or student notes that would invariably be available during those times when the student is tested in the world of intellectual performance, then what are we testing? And what lessons about adult standards and authority are learned if resources are arbitrarily withheld—arbitrarily in the sense that the student knows full well that during all genuine work situations, resources are available. Here we see most clearly that the use of tests, combined with secrecy, causes a moral imbalance that needs to be closely monitored.

Shades of gray exist with respect to possible access to resources. A physics teacher I know allows students to bring an index card to the final exam containing anything they want on it. Not only does this device perhaps sharpen and focus prior study, it may provide the teacher with an additional assessment: do the students know which facts, figures, and formulae are of most worth as resources? (An unintended outcome, of course, is that students learn to write in superhumanly small ways.) Some state exams, such as New York's earth science exam, go half way by standardizing the resources. They provide a booklet of tables and charts for the student to use during the exam.

Whatever the technical merits in forbidding the use of resources, another pedagogically harmful aspect of the prohibition is that what other people know (and which might be of help) must remain a secret in such a situation.
Our testing, wittingly or not, daily makes clear the dysfunctional lesson that intellectual assessment must be conducted in silence and in isolation from others. To ensure the validity of the inference that a person's score is truly theirs such a practice may well be defensible in many instances. But another kind of validity may thereby be sacrificed. Are we actually finding out if the most basic feature of adult work life—namely, that we can use all resources, including others, as necessary to get the job—is never assessed? Nor is it infeasible to build the effective seeking of advice into a test: many district writing assessments and performance science tests are now multiday, and build into the process the possibility of collaboration after an initial rough draft in isolation.

There is potentially a deeper moral harm to the student when possible colleagues are viewed as nothing more than possible accomplices and cheaters. Piaget (1932/1974) went so far as to suggest that exam-oriented schooling, where educators "have found themselves obliged to shut the child up in work that is strictly individual" reinforces all the child's most self-centered instincts and "seems contrary to the most obvious requirements of intellectual and moral development."

What, then, of the student's presumed right to ask questions of not just other test-takers but of the assessor? That right threatens traditional standardization with its overarching concern for comparability of conditions of administration. We know that tests that permit assessor/student interaction are easily corrupted unless there are strict protocols and guidelines by which adult judgment about intervention is informed. But we can turn the question around: how can the scores on complex and ill-structured test questions ever be adequately reliable if there is inherent ambiguity to the test questions—yet, the student cannot clarify the questions or double-check on the proposed solution (as any realistic setting would allow)?

All these situations call for testmakers to develop more sophisticated questions and dynamic/interactive tasks, with a protocol for test administration that lays out in some detail the permissible and impermissible interactions. This has already been done in IQ testing and much qualitative research (such as moral development interviews). And we should demand of test companies as well as teachers the use of certain questions and tasks that permit the use of resources that are specified in advance (such as textbook, dictionary, calculator, and one's notes)—questions that can be given out after a preliminary secure section, perhaps. In terms of managing this logistically, the students would procure the resources they brought and/or those provided by the tester, from another room during a break in anticipation of the more open part of the test. Then, there would be no compromise to the first part of the test.

Post-Test Security

As for post facto security, there can be little justification for it, especially at the local level. Although it is true that certain proprietary (and economic) interests might be threatened with such a policy, those concerns must be of secondary importance in almost all cases as it is surely detrimental to the student's (and teacher's) education for the test papers to remain a secret. Arguments that such a policy imposes an undue hardship on test companies do not seem very compelling when New York state, the International Baccalaureate, and the Advanced Placement programs make their tests and student papers public each year.

Being unable to inspect the test after the fact means that the value of the test—its formal validation—remains an inappropriate secret when students and teachers never see nontechnical (and nonintimidating) evidence and argument for the validity of the test in the first place. In the long run this can become an issue of educational equity and access: the student test-taker is then prevented from gaining not only the incentives but the insights that come from tests that are authentic and thus educative—where test tasks and scoring standards instruct the student about real-world tasks and standards.

The post facto test security also clearly contributes to the situation where our tests are routinely designed as isolated, one-event experiences—intellectually unauthentic and against the student's long-term educational interests. Under such conditions, the real meaning of each test score, and the significance of that score for the student, can be said to be inappropriately secret. The student receives what are, in effect, noncomparable scores over time (as opposed to a system of recurring assessments, scored in terms of continuous progress in reference to known, stable standards, as in the ACTFL foreign language proficiency guidelines or the new British and Australian longitudinal assessment frameworks). This practice is thus linked to the counterproductive nature of indirect secure testing: the inherent failure of such tests to provide powerful, rich, and useful feedback to each student as to the nature of their errors, the validated importance of such errors, and diagnostic help in how the errors might be eradicated.

The Formal Positions of the Professions on Testing Ethics

We look in vain for adequate guidance from the profession on secrecy in testing. The APA/NCTME/AERA Standards for Educational and Psychological Testing (hereafter referred to as the "APA Standards") offer no counsel on the morality of test security. In neither the technical or ethical standards do we find discussion, for example, of the test-taker's rights with respect to adequate advance knowledge and preparation. The principles under Standard
16 on “Protecting the Rights of Test Takers” have to do with such matters as informed consent and the right to privacy about scores (Millman & Greene, 1989; Wiggins, 1993). The only relevant standard is Standard 16.2 which states that “test users should provide test takers . . . with an appropriate explanation of test results and recommendations made on the basis of test results in a form they can understand.”

This “form they can understand” need not include, however, direct access to the answered and scored test. An intriguing historical change of mind that relates to access to one’s test has in fact occurred in the APA Standards for Ethics in research. In a recent paper, Haney and Madaus (1991) usefully point out that a key clause of the 1977 ethics guidelines was deleted in the most recent (1981) edition. In the 1981 and 1985 versions of the Ethical Principles, the following section was removed: “Persons examined have the right to know results, the interpretations made, and, where appropriate, the original data on which the final judgments were made.” Haney and Madaus (1991) also note in describing the changes in the APA Standards that

Strong vested interests within the APA demanded changes that favored their constituents at the expense of the test taker. . . . The profession has yet to come to grips with how the Standards can work fairly and equitably for developer, user, and test taker without an enforcement mechanism . . . other than the courts.

If we are going to have standards, they should give primacy to test takers, not test givers (just as in the APA “Ethical Principles in the Conduct of Research with Human Participants” research standards there are now clear protections against inappropriate and misleading research on subjects). Where, for example, under Standard 16 is a discussion on the students’ right to minimal or defensible “security”? Shouldn’t one be presumed to have the right of access to test papers after they have been scored? Where are the standards that provide students with the realistic opportunity to challenge a score or the aptness of the test questions?

It seems to me no coincidence that we continually and unthinkingly retain the use of blanket secrecy in dealing with minors. To put it charitably, we, of course, know better than they what is in their interest. To put it somewhat cynically, children will never likely be protected from excessive secrecy or any other practice that depends on moral inequality as they will always be without adequate political clout to ensure that their moral equality is codified into law. Thus, Walt Haney, who has been an “expert witness” in the longstanding lawsuit over whether the test companies are obligated to release individual completed and scored tests (as opposed to just scores), told me that in his opinion the original test disclosure law could only have been passed in the first place because the major class of litigants were medical students, not school children.

**Toward an Intellectual Bill of Rights**

Principles for the conduct of assessment that put the student’s rights on a moral par with the technical needs of psychometricians and the policy needs of school board and community members are thus long overdue. I know of only a handful of school districts that have testing and assessment policies. And policymakers, educators, classroom teachers and district administrators need significant help in “testing” the tests—help that takes the typical complaints beyond the merely personal to the principled.

More formal and powerful help could perhaps come from a genuinely disinterested national organization, set up to review tests and test-maker claims, modeled along the lines of Underwriters Laboratory or the Federal Trade Commission—an idea offered by Haney and Madaus (1991) and others as a more appropriate venue than the de facto one of courts of law: “[We call for] the establishment of an independent auditing agency that would, without vested interest, evaluate tests and testing programs that profoundly affect the lives of examinees.” Legislation may well be part of the solution. In a recent article on what should be done at the Federal level to prevent testing-related abuses, Feuer, Fulton and Morison (1993) made a similar pitch: Congress might “focus on various proposals to certify, regulate, oversee, or audit tests,” including the establishment of an oversight agency.

The authors also argue for Congress to “require or encourage school districts to develop and publish a testing policy” among other suggestions. Each school district ought to at the very least state the permissible uses of each morally problematic practices as test security, scoring work on a curve, the use of nonarticulated and generic tests, the failure to require consistent grading among teachers for the same work, and so on, as discussed in the rest of the book. These policies would do more than state the values of the institution; they would provide a local procedure for ensuring that assessment practices of both students and educators are more publicly scrutinized, discussed, justified, and improved. Regional service agencies and the federally funded educational labs and centers might also be asked to serve as a clearinghouse for both policy statements and training in how to fashion one. And a set of student-focused principles concerning assessment has been proposed by the New Zealand Department of Education (1989).

More than generic guidelines are needed. Each school system ought to have a Student Bill of (Assessment) Rights that protects the inherently vulnerable student from the harms that testing easily leads to. Here is my rough draft of such a set:
A DRAFT "ASSESSMENT BILL OF RIGHTS"

All students are entitled to—

1. worthwhile (engaging, educative, and "authentic") intellectual problems that are validated against worthy "real-world" intellectual problems, roles, and situations.

2. clear, apt, published, and consistently applied teacher criteria in grading work and published models of excellent work which exemplify standards—minimal secrecy in testing and grading.

3. ample opportunities to produce work they can be proud of; thus, ample opportunity in the curriculum and instruction to monitor, self-assess and self-correct their work.

4. assessment, not just tests: multiple and varied opportunities to display and document their achievement, and options in tests that allow them to play to their strengths.

5. the freedom, climate and oversight policies necessary to question grades and test practices without fear of retribution.

6. forms of testing that allow timely opportunities to explain answers or justify answers marked as wrong which students believe to be apt or correct.

7. genuine feedback: usable information on their strengths and weaknesses, and an accurate assessment of their long-term progress toward a set of exit-level standards framed in terms of essential tasks.

8. scoring/grading policies that provide incentives and opportunities for improving performance and seeing progress against exit-level and real-world standards.

The Bill of Rights would be supported by explicit audit policies and oversight procedures to ensure that these rights were protected.

I am sorry to report that this idea of a Bill of Rights has been attacked by more than a few teachers when I have offered it in workshops. Some have actually angrily called for a prior list of student responsibilities (although I do not recall such a list in our Constitution). But rights are not earned. Perhaps nothing better illustrates why these rights are deserving of formal protection, given the uneven balance of moral power in both the testing and teaching relationships as traditionally defined, than such a grumpy posture. The implicit hypocrisy in the position of these teachers is easily made explicit when one asks them whether they would be willing to endure a professional performance appraisal operating under the same conditions as student testing.

GRANT WIGGINS

CONCLUSION

Assessment, to be educative and fair, needs to be more a matter of principle and less a matter of good intentions, technical tricks, mere habit, or personality. Overarching criteria for "testing the test" in terms of secrecy are long overdue. At the very least, the student is entitled to the minimal secrecy necessary, and the right to see the score justified—the modern legal fair trial procedures and principles. To be autonomous (i.e., as the word's roots imply, self-regulating) the student must confront assessment procedures that are maximally transparent. The student must always be able to verify the aptness of the test, the score, and the assessor's use of secrecy. And we must know what that the assessor wants is of objective intellectual worth. Otherwise, we resort to merely guessing or calculating as to "what they want." In practice, this means that testing contexts should always make it possible for the student to self-assess and self-adjust before it is too late. This might even extend to building in a self-assessment of performance by the student as part of the test; a preponderance of inaccurate self-assessments might well cast doubt on the validity of the test.

The combination of high-stakes and large-scale secret tests, in which the only apparent judges are electronic, threaten intellectual integrity even further because they heap harmful suspicion on human judgment. Intellectual freedom is thus threatened by secure testing in the same way that it is threatened by political rule that emphasizes ritual and secrecy above the consent of the governed.

Unending testing breeds student insecurity (and teacher insecurity, when the multiple-choice test is externally designed). Something as simple as not getting back one's test papers can have harmful intellectual and moral effects. Paradoxically, according to Piaget (1932/1965), even if I respect the assessors, the authority figures, I am unlikely to effectively understand and uphold their standards if they are not demystified. Intellectual and moral standards are neither understood nor internalized if they remain "external to the subject's conscience," adhered to by force of the inequality of relationship only. Isn't this what we see and decry in students all the time? Isn't that just what we mean by talking about student thoughtless behavior in both senses of that word? Listen also to our students' language in talking about work turned back! "She doesn't like my writing" is a comment indicating that the subjective and hazy desires of authority are being taught, not clear standards. Correctness of answers and arguments are justified only because authority says so. Seemingly careless and thoughtless work is the inevitable result.
So is passivity: students learn by such mysterious and one-way assessment that they cannot reframe the question, challenge its premises, reject questions as inappropriate, nor propose a better way to prove their mastery. The moral and political harm is significant. Too many students learn that they should just “give them what they want” and to accept or acquiesce in bogus but authoritative judgments.

Excellence is not (and must never seem to be) about satisfying elders and their inexplicit and apparently contradictory tastes. Excellence must be seen as the meeting of known and justifiable demands—what (objectively) works.

We then come to understand a lesson that is morally as well as intellectually empowering: the judge, too, is bound to intelligent principles.

Ultimately we must recognize that there is a fundamental cultural inequality at work in sustaining this harmful overreliance on test security. Pervasive test security is not necessary; it is merely convenient. One’s place in the educational and professional hierarchy matters more than some stated concern for validity or feasibility. One certainly does not hear a national hue and cry for the design of secure tests by which we might glibly compare all colleges or professional workers. Much test security is required merely to enable the test to be uniform, simple, cheap, and widely used.

An alternative approach is clear, and practiced in various districts and countries other than our own. Return control of assessment to teachers or their representatives; make the assessment more credible, contextual, and fair. Develop districtwide and statewide task banks of exemplary assessment tasks, tests worth teaching to, and emulating in one’s design. Develop clear and validated district performance standards through a benchmarking and community discussion process. Develop oversight and audit functions to ensure the soundness of the system.

When a pool of questions and tasks has been developed that is large enough to prevent corruptive teaching, make them public, as if to say to teachers and students alike: “These represent the range of tasks you must be good at to graduate. Now that you know this, the responsibility becomes yours (just as it does for the athlete, drafting student, artist, and debater). We will coach you until you know how to self-assess and perform well enough that we become obsolete as ‘teacher’ and upholder of standards. Then you will have become, quite properly, our colleague.” No mystery or secret there; just as it should be and must be if our aim is to empower students and teachers, instead of merely checking up on them and keeping them in check. Maybe the best advice to testmakers was offered by two researchers on performance assessment twenty years ago (Fitzpatrick & Morrison, quoted in Finch, 1991): “The best solution to the problem of test security is to keep no secrets.”

NOTES

1. See the “Environmental Impact Statement” developed by Fairtest (National Center on Fair and Open Testing, Cambridge, MA) as an example of how such a process might work.

2. Note that tests are not valid or invalid; in other words, inferences about results on tests are valid or invalid. Our failure to keep this point in mind is both common and unfortunate.

3. Consider, by contrast, the recent British manifesto underlying their new national assessment design (which will rely heavily on classroom-based, teacher-overseen assessment): “The national system should employ tests of a wide range of modes of presentation, operation, and response— a mixture of tests, practical tasks and observations should be used in order to minimize curricular distortion... The Group has no doubt that it a successful system depends upon teachers’ confidence in it... the report recommends that teachers be given more support in assessment, especially by providing them with a wider range of diagnostic tests... the tests should be so unobtrusive as to seem to students no different than typical classroom activity.” From the Task Group on Assessment and Testing Report (TGAT), Department of Education & Science (1988).

4. The vaunted college-admissions test, the SAT, is a lovely example of what happens when secrecy shrouds the testing process. It is not an achievement test, but it is used to whip districts and states into a frenzy about school achievement—despite the explicit warning by ETS in their material not to do so, and the obvious fact that such tests depend more on socially constrained views of “intelligence” (and, thus, socioeconomic status). And who recalls that the SAT was technically developed as an “aptitude” test, for equity reasons—namely, to find students with potential in the hinterlands whose achievement might be limited by their local schooling?

5. Educators in other countries derisively refer to our fetish for increasingly relying on multiple-choice tests as “the American solution” to educational problems.

6. Many educators do not seem to realize that establishing “test validity” is not merely a technical matter based only on technical rules and criteria. Validity can only be established through argument to show that the test results really can be used to derive broader inferences about the value of the test scores vis-à-vis the wider world. Test makers need to justify their claims, appealing to empirical evidence, not just statistics, something few school officials and policymakers require.

7. The reader should not hear this in a utilitarian or mechanical sense. Artists often talk this way: indeed, the AP art Portfolio judges “quality” in reference to whether the student’s product “works in its own way” to accomplish the artist’s end.

REFERENCES


Authentic Assessment and Curriculum Integration: Natural Partners in Need of Thoughtful Policy

LESLIE C. SOODAK and GISELLE O. MARTIN-KNIEP

In this article, we examine the relationship between the curriculum and the assessment practices of a group of teachers to determine the degree to which change efforts, directed at teachers' curriculum and instructional approaches, have had an impact on other areas of their professional practice—more specifically, on the assessment strategies they use. We present findings from a study on the assessment strategies of 14 teachers who have participated in a comprehensive and long-term curriculum integration initiative. The findings indicate that teachers continue to rely on traditional assessment methods, although these are inconsistent with the goals of an integrated curriculum. These findings suggest that efforts to bring about educational change need to strengthen the linkages between curriculum, instruction, and assessment.

INTRODUCTION

Recent calls for the restructuring of American education have led to a barrage of reform initiatives in schools throughout the country. Whereas the overriding goal of these initiatives is to improve schools and raise academic

THOPO'S NOTE: This study was supported with funding from the American Forum for Global Education, the High School Division of the New York City Board of Education, and the Jaimiel Foundation. We would also like to acknowledge the support of Jaimiel Cloud,