RICHARD J. STIGGINS

Improving Assessment Where It Means the Most: In the Classroom

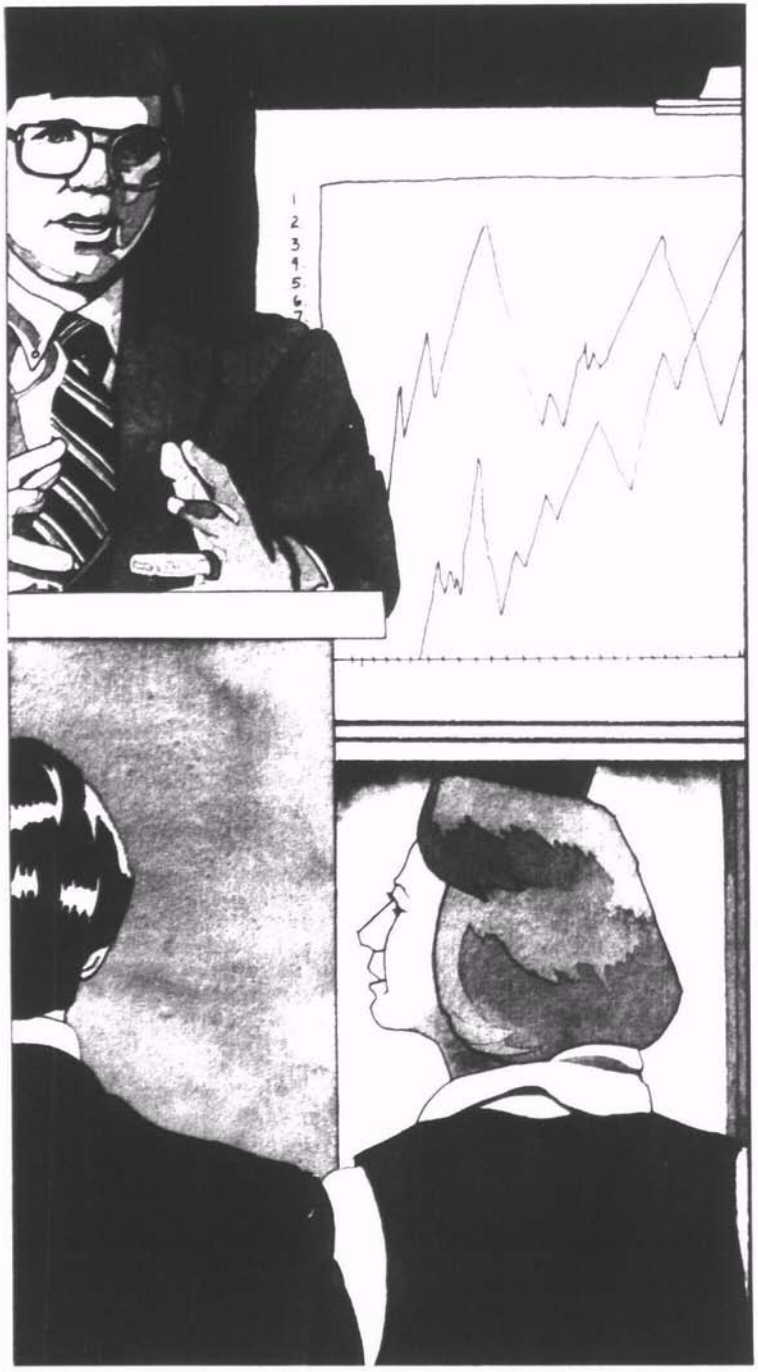
Despite the fuss over standardized testing, the information teachers need and use comes mostly from tests they develop themselves.

Mention achievement testing and many people think immediately of the Scholastic Aptitude Test, Iowa Tests of Basic Skills, California Achievement Tests, and the like. These are, after all, the test scores most often reported by the media and addressed in professional journals. In fact, if we were to try to define "assessment" *without looking into a classroom*, we could reasonably conclude that it is the measurement of students' achievement by standard multiple-choice test items for the purpose of showing the public whether or not schools are doing the job.

However, viewing assessment from *inside the classroom* produces a completely different picture. There, teachers use many kinds of assessments—some standardized and some individualized, some based on paper-and-pencil tests and some based on observation and judgment, some formal and some informal (Stiggins and Bridgeford, in press). Their assessment methods vary greatly depending on purpose, grade level, and subject matter.

Teachers face many kinds of assessment demands and measure dozens of student characteristics, from ability to achievement to personal traits and social characteristics. Not only must they chart these factors for individual students, but they must also maintain a sense of how the class as a whole is functioning. Teachers simply do not have the time to measure a student trait and then wait for several weeks while the computer transforms the result into a percentile or grade equivalent score. To make the decisions that keep instruction moving, teachers need data immediately, and they must gather them from a constantly changing environment. They rely most heavily on assessments provided as part of instructional materials and assessments they design and construct themselves—and very little on standardized tests or test scores. But more important, as virtually every recent study of school testing tells us, teachers depend heavily on their own observations and judgments—not just on paper-and-pencil tests (Herman and Dorr-Bremme, 1982; Kellaghan, Madaus, and Airasian, 1982; Salmon-Cox, 1982; Stiggins and Bridgeford, in press). Clearly, the assessments that influence classroom learning and students' academic and personal self-concept are those developed and used by teachers on a daily basis.

"Administrators must show that schools are doing the job they are supposed to do—produce learning."



Unfortunately, these teacher-developed assessments seldom attract publicity, have not been the subject of measurement or effective schools research, are rarely the focus of inservice training for teachers or administrators, and are usually conducted by teachers without technical support services from the building or district. The time has come to understand why the tests that mean so much to students and teachers in the classroom mean so little to the public, administrators, and reseachers—and, conversely, why the standardized tests that mean so little to the classroom teacher command so much attention from those outside the classroom.

Contributors to Divergent Perspectives

Over the decades, the measurement research community has tended to neglect classroom assessment in favor of research and development on large-scale assessment. As a result, our research on the classroom assessment environment is limited. Since most measurement textbooks are written by measurement researchers, they perpetuate the narrow view that assessment means paper-and-pencil testing. In the course of teachers' preservice measurement training (if they receive any, and many do not), they often learn that standard paper-and-pencil "objective" tests are the best means of measurement. As they begin teaching, however, they discover three information sources: the standardized testing program, tests provided in textbooks, and assessments they develop on their own. They quickly discover that just two of these sources are readily available for daily use, reflect daily instructional priorities in their classrooms, and produce results almost immediately. Since their own assessments (tests and observations) and text-embedded tests serve their information needs more effectively than do standardized tests, these are the tests teachers trust and value. With experience, they also come to trust their own observations and professional judgments regarding student achievement and rely on student behavior and

products as indices of growth and development.

Educational administrators face different decisions, have other information needs, and therefore have different values. Administrators perform a public relations function between school and community. They must show that schools are doing the job they are supposed to do—produce learning. That means they must report consistent, scientific, and objective test scores to the public. And because we are a competitive society both on the playing field and in the classroom, reported scores must offer comparisons among schools and districts. Teacher-made tests are not seen as satisfying these criteria because they vary greatly from class to class and are hard to compare. In addition, their quality is sometimes questioned. These factors combine to make standardized tests more attractive to administrators.

Unfortunately, the accountability demands placed on administrators give rise to a critical role conflict for the administrator. As *accountability agents*, they must give much time and energy to the large-scale standardized testing program. As *instructional leaders*, administrators have a duty to support the classroom assessment efforts of teachers. Two factors influence which role wins out: resources and knowledge. The amount of time, energy, and money administrators can give to assessment issues is severely limited, and few administrators have the training or experience to help teachers with daily assessment design, use, or quality control. Thus, they devote their limited time and energy to the standardized testing program, and teachers are left on their own.

This preoccupation with accountability through visible test scores reinforces in the public mind a very narrow conception of the intended outcomes of education and a restricted understanding of what kinds of assessment are "acceptable" indicators of achievement. Repeated reporting of standardized test results under the guise of "keeping the public informed" reinforces the incorrect im-

"As *instructional leaders*, administrators [also] have a duty to support the classroom assessment efforts of teachers."

pression that we can reflect the very complex student characteristics we seek to develop in a single set of paper-and-pencil test items yielding a single score. Perhaps even more sadly, it also reinforces the impression that the classroom assessments and professional judgment of teachers, while acceptable for daily instructional decision making, are not acceptable for public consumption.

The public has also played a role in the development of current testing values. Parents, taxpayers, and board members have remained on the surface of school testing issues, asking only the most superficial questions. The final measure of our educational progress seems to rest with one question: are SAT scores going up or down? Never mind the fact that such scores fail to reflect anywhere near the total range of our intended educational outcomes, or that the sample of students tested is in no way representative of the total student population. Simplicity is the watchword for critics of education, despite the probability that the simplicity of test scores leads to a misrepresentation of the outcomes of American education.

The Impact of Value Differences

These value differences are having a decidedly negative effect on school improvement. To illustrate, the past decade has seen the rapid accumulation of studies of effective school practices. In virtually every case, the criterion variable used to identify those practices has been an increase in standardized test scores—despite the fact that we know these tests represent only a few of the outcomes we want students to achieve. The day-to-day measurement of student growth has played virtually no role in this research, even though these are the measures that tell teachers which practices do or do not work.

Our current assessment values may also be contributing to inadequate daily assessment of student achievement in some classrooms. Since we have rarely inquired into the quality of teacher-developed tests, offered training in classroom assessment, or included classroom assessment in the principal's leadership role, we simply do not know how well teachers measure student achievement or how to help them if they need help.

To avoid these problems, we need to revise our research priorities, improve the relevance of educator training in assessment, and include technical assistance in assessment as a key component of the administrator's leadership role.

Research Priorities

Measurement researchers might enlarge their understanding of school testing by studying the relative impact on student learning of teacher-made versus standardized tests. They might structure their research toward exploring the quality (reliability and validity) of teacher-made assessments, the role of teachers' judgment in assessing student capabilities and achievement, and the student's perspective—how assessment influences study habits, aspirations, and sense of control over personal achievement.

School effectiveness researchers, who have tended to rely on standardized test scores for judging effective practices, might broaden their criteria to include teachers' achievement measures. They could ask teachers with a long-standing record of outstanding classroom performance to describe instructional strategies that work for them in specific contexts over the long haul, and also ask them to provide assessment evidence of their own to prove that those strategies produce learning. Researchers might then screen the teachers' reports, evaluate the quality of the evidence, and synthesize the effective practices. Whatever common themes emerge could represent classroom practices capable of producing more than just higher standardized test scores. This kind of applied research could be conducted and used by instructional leaders at the school building level.

Assessment Training Priorities

The insights provided by such research might yield new ideas for training educators in educational assessment and its relationship to instruction. In the meantime, some assessment training priorities are already indicated for administrators, teachers, legislators, and the public.

District- and building-level administrators who lack expertise in assessment may wish to:

- Seek inservice training in standardized testing that includes a focus on its limitations.
- Gain the knowledge needed to guide teachers in the simple procedures of test design, construction, and quality control.
- Learn to help teachers use observations and professional judgments as valid and reliable indicators of student achievement.
- Learn to apply the basic principles of survey sampling methodology to districtwide testing programs to save money (Stiggins, 1975).

In short, administrators should seek the kinds of training and experience that would allow them to assist teachers with daily classroom assessment. Training guides by Spandel (1981) on classroom writing assessment and Stiggins (1984) on classroom performance assessment illustrate the kind of training needed.

Teachers may also need relevant, focused training in assessment. Preservice teacher training often fails to include a course on testing (Coffman, 1983), and assessment is rarely the topic of inservice training. As a result, teachers report that they learn how to measure student achievement from their colleagues and their own experiences as students; they rarely cite formal training as a source of ideas (Stiggins and Bridgeford, 1982).

The content for classroom assessment courses is suggested by a study of over 300 teacher-made paper-and-pencil tests conducted by the Cleveland Public Schools (Fleming and Chambers, 1983). In that case, teachers appeared to need training in how to (1) plan and write longer tests, (2) write unambiguous paper-and-pencil test items, and (3) measure skills beyond recall of facts.

Another study revealed that many teachers overlook key quality control factors in using performance assessment (observation and judgment) in the classroom (Stiggins and Bridgeford, in press). For instance, less than half of the performance assessments conducted by the teachers studied included written performance criteria, information provided to students on performance criteria, or multiple observations before making judgments (to ensure dependable results). Rarely did teachers plan scoring procedures in advance or define levels of performance. They seldom rated student performance without knowledge of the student's identity (to control bias or halo effects), cross-checked judgments with other performance indicators such as test scores (as a validity check), or relied on other teachers as raters of student performance (as a reliability check). In addition, teachers tended to rely on mental rather than written recordkeeping to store information regarding observed student performance.

This research suggests that training for teachers should include strategies for clarifying performance criteria, developing performance exercises, observing behavior, and evaluating it dependably. Such assessments play a key role in the measurement of basic communication skills, the performing and visual arts, psychomotor development, and many other skill areas. If assessment of these skills is of high quality, instruction is more likely to be of high quality.

Both teachers and administrators need training in specific strategies for:

- Clarifying the demands of various assessment purposes.
- Selecting assessment methods to match purposes.
- Designing or planning assessments that work.
- Ensuring the quality of those assessments.
- Building specific paper-and-pencil tests.
- Constructing assessments based on performance judgment.
- Measuring specific basic communication skills (for example, in reading and writing).
- Using modern technology in classroom assessment.
- Promoting testwiseness among students.
- Measuring higher-order thinking skills.

Legislators, taxpayers, parents, and the general public might also be trained to understand the limitations of large-scale testing programs—to understand that the mere presence of a testing program does not ensure quality education and more testing will not of itself produce better schools. The public must become aware of the full range of (a) complex student characteristics (not just mathematics or reading skills) that can be influenced by quality education, and (b) available alternatives for equitably measuring those characteristics. We need to develop a new generation of critical data consumers—a generation that knows the attributes of good assessment. Such training has been promoted in the past by the National Institute of Education through publication of guides like *Training Citizen Groups on Educational Testing Issues: A Trainers Manual* (Anderson, Stiggins, and Gordon, 1980) and *Your Child and Testing* (Herndon, 1982).

Ideally, one outcome of teacher and administrator training efforts on *classroom* assessment will be the increased ability of school leaders to balance their roles of accountability agent and instructional leader. By balancing these roles effectively, administrators can help ensure that:

1. *Communication channels will open to let teachers and administrators discuss issues that go beyond standardized test scores.* Teachers will be regarded as credible sources of information on student achievement and school effectiveness. In periodic news conferences, highly qualified master teachers might report on the status of schools based on their own sources of achievement data. These reports could be supplemented by performance-based data, such as assessment of writing samples read and judged by quali-

"Teachers simply do not have the luxury of measuring a student trait and then waiting for several weeks while the computer transforms the results into a percentile or grade equivalent score."

"We need to develop a new generation of critical data consumers—a generation that knows the attributes of good assessment."

fied teachers. Further, as part of the districtwide testing program, teachers could be asked their opinions on what proportion of their students are working up to full potential.

2. *Resources will be appropriated so that teachers can collaborate on assessment projects.* Discussion of assessment issues, for example, could lead to a common set of performance standards and an integrated program of instructions to meet those shared expectations. In addition, teachers might design and use tests together, observe each other during performance assessment, rate the performance of each other's students, evaluate one another's tests, or just share good ideas.

3. *Districtwide testing programs will be planned with an eye toward matching the cost of the program to its impact on student learning or just improving its impact and usefulness to those in teaching and supervisory positions.* Districtwide testing programs are typically intended to provide information for public consumption—for accountability. They represent only a rarely used secondary source of information to teachers. Therefore, in many districts it may be neither necessary nor cost-efficient to test every pupil. The savings that result from random sampling could be used to fund inservice programs to develop teachers' test preparation skills and promote understanding of assessment issues.

4. *Resources will be allocated to provide technical assistance to teachers in the development and use of classroom assessments.* Expertise might be provided to help teachers design and construct better tests, score them, and analyze results. Individualized help might be made available for teachers who wish to try some new and better way to measure outcomes that are important to them. Such personalized, continuous training could have a very positive impact on instruction—not only because teachers would develop or improve testing skills, but also because the provision of specialized training would say to teachers clearly and unequivocally that their assessment insights and judgments are valued and respected.□

References

Anderson, B.; Stiggins, R.; and Gordon, D. *Training Citizen Groups on Educational Testing Issues: A Trainer's Manual.* Portland, Oreg.: Northwest Regional Educational Laboratory, 1980.

Coffman, W. E. "Testing in the Schools: A Historical Perspective." Paper presented at the UCLA Center for the Study of Evaluation Conference on "Paths to Excellence: Testing and Technology, Los Angeles, 1983.

Fleming, M., and Chambers, B. "Teacher-Made Tests: Windows on the Classroom." In *Testing in the Schools.* New Directions for Testing and Measurement, No. 19. Edited by W. E. Hathaway. San Francisco: Jossey-Bass, 1983.

Herman, J., and Dorr-Bremme, D. R. "Assessing Students: Teachers' Routine Practices and Reasoning." Paper presented at the annual meeting of the American Educational Research Association, New York, 1982.

Herndon, Enid. *Your Child and Testing.* Washington, D.C.: National Institute of Education, 1982.

Kellaghan, T.; Madaus, G. F.; and Airasian, P. W. *The Effects of Standardized Testing.* Boston: Kluwer-Mijhoff, 1982.

Salmon-Cox, L. "Teachers and Standardized Tests: What's Really Happening?" *Phi Delta Kappan* 63 (1982): 631–634.

Spandel, V. *Classroom Application of Writing Assessment: A Teacher's Handbook.* Portland, Oreg.: Northwest Regional Educational Laboratory, 1981.

Stiggins, R. J. "Alternatives to Blanket Standardized Testing." *The Journal of the National Education Association* (March-April 1975): 38–40.

Stiggins, R. J. *Evaluating Students by Classroom Observation: Watching Students Grow.* Washington, D.C.: National Education Association, 1984.

Stiggins, R. J., and Bridgeford, N. J. "Final Research Report to the National Institute of Education on the Nature, Role, and Quality of Classroom Performance Assessment." Portland, Oreg.: Northwest Regional Educational Laboratory, 1982.

Stiggins, R. J., and Bridgeford, N. J. "The Ecology of Classroom Assessment." *Journal of Educational Measurement*, in press.

Richard J. Stiggins is Director, Center for Performance Assessment, Northwest Regional Educational Laboratory, 300 S.W. Sixth Avenue, Portland, Oregon 97204.