Matrix Sampling

Testing sample groups of students in a variety of subject areas can give schools a good estimate of the overall quality of their learning programs.

HOWARD EBMEIER

Without a systematic method of monitoring processes and products, it is extremely difficult to determine the overall effectiveness of a school program or identify areas that need attention. Indeed, the school effectiveness studies have found that more effective schools depend on assessment systems that use criterion-referenced or standardized tests to gather student achievement data.[1] The managerial problems associated with such instruments are considerably less than with other assessment techniques, such as follow-up studies or classroom observations, and they may be the only reliable instruments available to a district.

Regardless of which instruments are used, however, the mass testing of students for program evaluation poses problems. While mass testing continues to flourish in elementary schools, it has generally been abandoned at the secondary level. This is chiefly due to student motivational problems inherent in lengthy testing sessions, the great diversity of courses offered in junior and senior high schools, and the procedural problems caused by mass testing.

To overcome these difficulties, the Community Unit School District in Wheaton, Illinois, has employed a matrix sampling technique. Once a year in a two-hour period, we are able to gather evaluative data in at least ten content areas by administering conventional standardized and criterion-referenced tests to sample groups of students.

Matrix sampling is the in-depth analysis of a subset of all possibilities in order to make inferences about all of the parts. Industry frequently employs sampling techniques to maintain quality control. Car Company A, for instance, might pull every tenth engine for a detailed inspection, or Rubber Company B might check the specifications on every 100th tire to see if it conforms to guidelines. By using such sampling procedures, these companies obtain a good estimate of the overall quality of their product without examining each one—an expense they could hardly afford.

Step 1: What's Important When?

We employ this same technique to gather student achievement, attitude, and performance data in a more efficient manner than if all students took the same tests. We first establish in what areas and at what grade levels students are expected to have mastered given content. Detailed behavioral objectives could be used, but we choose to select rather broad areas. For instance, in the example in Figure 1, mastery of library skills is an expected outcome in the seventh to ninth grades, while American government skills are not taught until the eleventh and twelfth grades. In contrast, language arts and physical education are emphasized at all grade levels.

After mapping out the content, we determine how many samples to draw. A good rule of thumb is to select about one-third of the total population for each group, or three groups per grade level. This sample size is reasonably large but ensures that the average score of the sample will be within three percentage points of the value obtained if everyone took the test.[2]

Three tests can be given if each sample group takes only one test. But if each group takes three tests, nine different tests could be administered per grade level. At 40 minutes per test, the total testing time is only two hours—a vast advantage over mass testing systems that would require about six hours to administer all nine tests to every single student.

Step 2: What Do We Want to Know?

In the next step we determine in which areas we want to obtain information. Local circumstances play a large role in this decision; indeed, this is one attractive feature of matrix sampling. Testing programs can be custom-tailored to district needs. In our district, the three sample groups at each grade level each take two tests. The circled areas in Figure 1 reflect the areas and grade levels at which tests are administered each year. We assess student math performance every year because it is a critical basic skill, and examine spelling skills at only the ninth and eleventh grades.

Westinghouse

Howard Ebmeier is Director of Research and Evaluation, Community Unit School District 200, Wheaton, Illinois.

Figure 1. Expected Student Competence and Areas of Content Coverage Across Grade Levels

	Content										
Grade Level	Library Skills	Spelling	PE	Math	Language Arts	Reading	Biology	Physical Science	American Government	Health	Attitude Toward School
7	⊗	x	⊗	⊗	⊗	⊗					⊗
8	⊗	x	⊗	⊗	x	⊗	⊗	x			x
9	x	⊗	x	⊗	⊗	⊗				⊗	⊗
10		x	⊗	⊗	⊗	⊗	⊗				⊗
11		⊗	x	⊗	x	⊗		⊗	⊗	⊗	x
12		x	⊗	⊗	⊗	⊗	⊗	x	x		⊗

○ Areas to be tested, assuming three samples per grade level and two tests per sample.
x Indicates that a particular subject is emphasized or taught at that grade level.

This pattern could change as a result of the evaluative data obtained. For instance, we may have discovered that the math program is strong across all grade levels but a deficiency exists in spelling at the two grade levels tested. The following year we could reduce the number of grade levels given the math test and increase the number of spelling tests. We might also give a diagnostic spelling test to isolate the cause of the spelling problems. Another option is to rotate content areas tested, especially those not typically considered basic skills. Thus, one year a consumer education test could be given and in other years health, music, art, or mechanical skills could be assessed. In this way, over a five-year period, evaluative data are gathered in almost every program area.

Step 3: How Are We Going to Find Out?

In step 3, we select testing instruments. Involvement of subject area specialists or department chairpersons at this point is critical to ensure that the tests we select reflect the taught curriculum. The involvement of these individuals also ensures their commitment to the matrix sampling concept. Without such a commitment and a belief in the usefulness of the data being collected, little improvement at the building level can be expected as a direct result of the testing program.

Whether to use a published test or to construct a new instrument depends on the degree of overlap between the test and the curricular program. It may be better to use an existing test that reasonably covers the real (taught) curriculum, since developing and validating a new test is costly and time-consuming. The key factor in this decision is the degree to which the subject area specialists or department chairpersons believe the test covers important content that all students should know and accurately measures the school and department goals in that particular area.

Step 4: Finding Out

The remaining steps are mostly procedural but required logistical planning. They include the assignment of students and teachers to testing sections and the actual distribution of material. Students can be assigned to sections (samples) quite easily by matching students' names to random numbers. In our first two years using this system we employed a secretary for this task (one day) but have since computerized the operation. Either method is relatively quick and straightforward. The principal then assigns each section a room location and proctor. All teachers should be involved in administering the tests, and the number of students per section should not exceed 30. Having all the teachers participate builds commitment, and limiting section size increases student accountability and reduces motivational problems commonly associated with testing in large auditoriums or gyms.

Regardless of the care exercised in the development of a matrix sampling system and the efficiency of the student sampling and section assignment, if the packaging and delivery of the various testing materials to the teacher proctors are deficient, the whole process will be impaired and could fail. It is, therefore, important to clearly communicate expectations and arrange the testing materials and instructions in the most understandable way possible. We condense the instructions into a one-page format; bundle together the tests, answer sheets, instructions, and return envelopes; and place the section number, room number, proctor's name, number of students, names of the tests to be administered, and time limits on the outside of the bundle. This makes distribution easy and serves as a convenient materials list.

As the testing time approaches we communicate to the student body the procedure of the testing program and the importance of doing well. To increase motivation we record students' scores on their transcripts via gum labels and return the results to them as soon as possible (usually within three weeks). Given that the students know these records are available to colleges and potential employers, that the total testing time will be relatively short, and that proctors will be closely monitoring their behavior, we encounter few motivational problems.

Answer sheets can either be scored by hand via district resources, or sent to the test publishers. The only remaining task, but probably the most difficult, is to use the results wisely. We provide an inservice program that describes the results and shows teachers how to use the data.

This matrix sample testing program enables us to collect a large quantity of information about a broad range of student attitudes, achievement levels, and physical development skills with minimal testing time, cost, and student motivational problems. With the objective data derived from the matrix, we can identify our strengths and weaknesses and, most important, improve the quality of the overall curriculum.□

[1]For further information, see Ronald Edmonds, "Some Schools Work and More Can," *Social Policy* 9 (1979): 32; W. B. Brookover and others, *School Social Systems and Student Achievement: Schools Can Make a Difference* (New York: Praeger Publishers, 1979); and *Educational Leadership* 37 (October 1979), theme issue on effective schools and effective teaching.

[2]M. Allen and W. Yen, *Introduction to Measurement Theory* (Monterey, Calif.: Brooks/Cole Publishing Company, 1979).