

thinkpoint

What does it mean to use matrix sampling in student assessment?

by Edward Roeber, Ph.D.

INTRODUCTION

In most state- or district-level student assessments, all students are assessed on the same content using the same set of test items. Doing so helps to assure that content considered important is used in the assessments given to all students. The advantage of this approach is that the results can readily be compared across students, classrooms, schools, and districts, since all students have responded to the same items. Because the number of students who took these common tests can range from a few students to many, it may also be possible for educators to examine the performance of students on each test item or set of test items to make comparisons across items or content standards (such as, “on which item or set of items did students score the highest or the lowest?”).

There are times, however, when testing all students on the same set of items may not be a priority or even necessary. These include times when new items are being field tested, or times when only group-level performance is desired. In these instances, every-pupil testing on a common set of items

can be viewed as an inefficient and unnecessary use of testing time.

Recent concerns about students spending too much time in testing

have led some to ask the question: “Do we need to administer all of the test items to all of the students?”

The answer to that question is: “It depends.”

The most positive impact of matrix sampling is the reduction in the total amount of time needed for testing while still obtaining group-level estimates of student performance. This can save assessment administration time and scoring time, and reduce the costs of assessment without adding much to the data analysis and reporting tasks.

What is matrix sampling?

Sampling described here is the purposeful selection of a subset of test items or persons. To be truly useful when a sample is selected, each test item or person needs to have an equal chance of being selected. When this occurs, we call it random sampling. Here, the word “random” means “equal chance,” not “haphazard” or “uncontrolled.”

The key is to draw representative samples—of items or of students.

Matrix sampling is the selection of both things (i.e. test items) and people (i.e., students). The first step is to construct a matrix containing all test items and all students. Step two is to choose from among two basic models of matrix sampling:

Matrix sampling model 1

In this matrix-sampling model, the overall assessment is divided up into several parts. All students take part in the assessment, but each student takes only one part of the overall assessment. The number of parts into which an overall assessment is divided is often determined by the number of items and the number of students; the key is to assure as large a sample as possible of students who take each part of the overall assessment while striving to reduce testing time per student as much as possible. In this model, individual items or sets of items are randomly assigned to each part.

Example — A 120-item test is to be administered to 600 students. A sample size of 150 students per item (sufficient for showing district-level student achievement) might mean dividing the 120 items into four parts, each containing 30 items. Each student would take one-fourth of the test forms (either test 1, 2, 3, or 4). If a more precise estimate of student performance is desired, the 120-item test could have been divided into only three

parts, each containing 40 items, which is still a reduction in the number of test items any student takes, just not so much as seen with the 30 item test.

Matrix sampling model 2

In this model, a carefully selected sub-set of items from the entire test is administered to all students. The subset of items selected should strongly predict performance on the entire test. For example, select about a third of the items from an entire test that

most strongly predict overall test performance. In order to do this, a representative group of students needs to take the entire test and then statistical techniques are used to select the items that best predict overall test performance.

Example — A 45-item test might be administered to all students; then a statistical sample is selected that includes about 1/3 (15) of the items that predict well overall performance on the full test. In the future, all students would take the same 15-item test. (See Figure 2.)

Figure 1: Example of a four-part test given to five students

STUDENT	TEST PART 1	TEST PART 2	TEST PART 3	TEST PART 4
1	x			
2			x	
3		x		
4				x
5			x	

Figure 2: Example of a sub-test of 15 items randomly selected from a 45-item test

STUDENT	ITEM 4	ITEM 7	ITEM 8	ITEM 11	ITEM 15	ITEM 21	ITEM 22	ITEM 27	ITEM 35	ITEM 38	ITEM 39	ITEM 41	ITEM 42	ITEM 44	ITEM 45
1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

How and when might matrix sampling be used?

Matrix sampling is used in several ways in large-scale assessment programs, especially when individual student reporting either is unnecessary or of much less importance. Many assessment programs use matrix sampling for embedded field testing of new test items within their operational tests. The advantage of embedded field testing is that each student only has to take a few of the field test items, the items are included within the actual test students are taking so that student motivation is the same (since they don't know which items are actual or field test ones). Also, the time necessary to obtain student test performance can be minimized, especially when many students take the operational test, since samples of 500-1000 can be sufficient for

field testing purposes.

Matrix sampling is also used when only an overall estimate of student performance is necessary or desired. Since detailed information on student skills is not desired, it is possible to use either Model 1 or Model 2 to obtain overall test performance. An illustration of this is national tests such as the National Assessment of Educational Progress (NAEP) that report test performance only at the state and national levels, not that of individual students, schools, or districts.

Finally, educators who have large classes and simply want to report class-level data may also employ matrix sampling. For example, if an arts educator wishes to report the achievement of the overall band or orchestra (comprising 100 students) on a performance assessment of

playing capabilities (which can be time-consuming to administer and score), the teacher might select either a subset of students or divide the assessment into two or more parts and give each student only one of the several parts. This classroom application of Model 1 would provide useful information at the group level.



When might matrix sampling not be good to use?

It is not wise to use matrix sampling when an educator needs a detailed profile of the performance of a student. Even when only class-level performance is desired, educators must ask: "Are there enough students on which to divide up the test and still have adequate numbers of students in each group on which to report group results?" since using Model 1, each student will be assessed on only one portion of the standards. In a large school or district, with many students, it may be possible to use matrix sampling, so long as it is not necessary to show such information for each student.

Model 2 may not work well at either the student or the school/district level, since students are assessed only on a portion of the overall set of content standards (and accompanying test items). No information would be available on student performance on the test items not used, which could especially be an issue if the assessment has multiple strands of content and representing each strand is desired. Thus, this model may not be suitable for *those desiring detailed item information* from all students on all strands and content standards.

What are the impacts of using matrix sampling?

The most positive impact of matrix sampling is the reduction in the total amount of time needed for testing while still obtaining group-level estimates of student performance. This can save assessment administration time and scoring time, and reduce the costs of assessment without adding much to the data analysis and reporting tasks. If group-level data is all that is needed, then matrix sampling could be an attractive option for an assessment program.

Matrix sampling, if not carefully considered and implemented, may not yield the desired type or quality of assessment information. If group sizes are too small, or the pool of items is divided into too many sub-tests, the group results may not provide a good estimate of the performance of all students. Assessments based on only a portion of the content standards may not yield the desired comprehensive assessment information.

These pros and cons need to be

thoughtfully reviewed before a decision to use matrix sampling is made. Carefully implemented, this strategy can help to make assessment more useful and less burdensome. It can help achieve overall state goals for assessment and accountability without sacrificing quality assessment, thus permitting other types of assessment—those used to actually help students learn—to be used without increasing the overall burden of assessment on students.

Author

ED ROEBER, PH.D., is the assessment director for Michigan Assessment Consortium (MAC) projects. He co-chairs the MAC's Assessment Resource Development Committee and serves as Board Secretary/Executive Committee. Dr. Roeber has earned degrees in Psychology and Political Science (B.A.) and Educational Psychology/Measurement and Evaluation (A.M. and Ph.D.) from the University of Michigan.

Figure 3: Pros and cons of using matrix sampling

Consider using matrix sampling when...	<ul style="list-style-type: none"> ■ Individual student reporting is not required ■ Field testing new test items within an operational test ■ Estimates of student performance, by group, are all that are required ■ Reporting class-level data for large classes
It is not wise to use matrix sampling when...	<ul style="list-style-type: none"> ■ Educators need a detailed profile of student performance ■ Educators need detailed item information from all students on multiple strands and content standards
Potential benefits of matrix sampling...	<ul style="list-style-type: none"> ■ Reduction of total time needed for testing ■ Reduction of the costs of assessment
Potential drawbacks from using matrix sampling	<ul style="list-style-type: none"> ■ Wrong type of assessment information (e.g. assessments based on only a portion of content standards may not meet desired level of comprehensive information) ■ Lower quality assessment information (e.g. if group sizes are too small, or item pool is too subdivided, group results might provide inaccurate estimates)

To learn more

Further reading about models of public assurance that seek to rely less on large scale standardized measures:

Childs, Ruth. (2003). Matrix Sampling of Items in Large-Scale Assessments. Toronto, ON: Ontario Institute for Studies in Education of the University of Toronto.

pareonline.net/getvn.asp?v=8&n=16

Information on how the National Assessment of Educational Progress (NAEP) uses sampling

nces.ed.gov/nationsreportcard/about/samplesfaq.aspx

nces.ed.gov/nationsreportcard/tdw/sample_design

nces.ed.gov/nationsreportcard/about/nathow.asp

www.nagb.org/content/nagb/assets/documents/naep/kolstad-sampling.ppt