Testing the GED

In his article on the high-school graduation rate (“Tassels on the Cheap,” Feature, Fall 2002), Duncan Chaplin implies that the General Educational Development (GED) tests represent a lower academic hurdle than graduating from high school. He is simply wrong.

To earn a GED, candidates must pass five separate tests covering math, science, reading, writing, and social studies. The score necessary to pass does not reflect a minimum level of academic achievement; in fact, the passing standard is now set so that 40 percent of all high-school graduates would fail the test.

The GED is designed to document whether a student has mastered a level of knowledge roughly comparable to that of high-school graduates. It makes no promises about lifetime earnings, participation in postsecondary education, or job satisfaction. No test would ever make such promises. But what it purports to do, the GED does very well.

The GED Testing Service (a program of the American Council on Education) strongly encourages students to stay in school and graduate. For many, however, this may not be an option. For them, the GED tests offer an academically rigorous second chance.

MICHEL BAER
American Council on Education
Washington, D.C.

Duncan Chaplin responds: I agree that the GED provides an important second chance to adults without a regular degree. However, equating the GED with a regular degree does a great disservice to teenagers, since GED recipients are far less likely to complete college and earn far less on the labor market than those with regular degrees.

Even Baer’s claim that GED recipients have similar academic skills is highly suspect. It is true that a random sample of high-school seniors who were likely to graduate were given a GED, and 40 percent did not pass. However, unlike actual GED recipients, none of this random sample had an incentive to pass, took it multiple times, or prepared for the test. The fact that many did not pass is hardly surprising.

Margin of error

Dale Ballou (“Sizing Up Test Scores,” Forum, Summer 2002) sounds legitimate warnings about some potential problems associated with using value-added assessment as a way of evaluating schools and teachers. However, if appropriate analytical strategies are used, value-added assessment need not pose the dire risks that his article implies.

First, Ballou overstates the risk to value-added assessment that comes from the question of whether or not scales can be developed with equal intervals. There is no unanimity of opinion on this subject within the psychometric community, as the article implies. There is one school of thought that says equal-unit scales can be developed and used. However, whether a “pure and perfect” equal-unit scale exists is not the critical question. The question should instead be, “If scales from a testing regime are used within a value-added process, is there evidence that measures of student progress are influenced by the distribution of student achievement levels in schools or classrooms because of a lack of equal-interval scales?” After appropriate analytical investigation, if evidence arises that measures of student progress are indeed influenced by positions on the scales, then statistical accommodation would be required within the value-added modeling process. I agree with Ballou’s warning. However, I suggest that, if the empirical evidence points to a problem, there are appropriate analytical solutions to the problem.

Those who do not possess some basic statistical knowledge can easily misconstrue the part of the article that asserts that measures of gain innately have more noise than raw scores. A very basic statistical fact is: the variance of a difference between two variables is equal to the sum of the variances minus two times the covariance between the two variables. This is why, in our modeling efforts, we do massive multivariate, longitudinal analyses in order to exploit the covariance structure of student data over grades and subjects to dampen the errors of measurement in individual student test scores. This leads to an improved ability to distinguish the effectiveness of various schools and teachers. However, even with the most sophisticated statistical approaches, for some subjects and grades the ability to differentiate among educators may allow only a small number of schools and teachers to be identified as being different from the average educator. This should not be considered an obstacle to the deployment of value-added measures as a component of teacher and school evaluations, but as an attribute that allows the data to be used consistent with the reliability of the indicators that can be gleaned from them.

Ballou’s cautions are appropriate. However, they do not lead to the judgment that value-added assessment for purposes of evaluating teachers and schools is invalid.

WILLIAM L. SANDERS
SAS Institute, Inc.
Cary, N.C.
Dale Ballou responds: Responses to my article on value-added assessment suggest that it may be useful to put my arguments in a broader context.

Because test scores are noisy measures of true achievement, the quality of the information in these scores varies from teacher to teacher. In a letter published in the Fall 2002 issue, J.E. Stone alludes to ways of minimizing noise, but those statistical adjustments, which are used in the most sophisticated value-added assessments, have the consequences I described in my article: a teacher’s rating will depend on the subject she teaches, how large her classes are, how long she has been in the current system—in short, any factors that affect the quality and quantity of the available data. Some outcomes will be difficult for teachers to understand. For example, the test scores of Mr. Jones’s students may rise more than the test scores of Ms. Smith’s, but Ms. Smith, not Mr. Jones, may be recognized for excellence—because the state possesses better information about Ms. Smith. As long as the purpose of the system is purely diagnostic, these disparities are unlikely to cause much dissatisfaction. But when rewards and sanctions are attached, they will probably become a major issue.

I also expressed concern that test-score scales lack an essential property for comparing test-score gains: an interval of, say, 50 points at one point on the scale cannot be assumed to represent the same achievement growth as a 50-point interval elsewhere on the scale. Stone does not say much about this other than that my criticisms are “based principally on theory, not evident deficiencies.” On this point he is right, but only because these deficiencies will not be apparent until a single test is scaled multiple ways and the resulting assessments of schools and teachers are compared. Because practitioners of value-added assessment never do this, they cannot take comfort in the absence of “evident deficiencies.” William Sanders (see above) writes that psychometricians are not all of the view that achievement scales lack an equal-interval property. This is so. Sanders also writes that statistical adjustments can accommodate any problems that might arise from scale indeterminacy. On this point, Sanders and I do not agree. Once a scaling decision is made, it affects the numbers on which all subsequent statistical analyses are based. It cannot be adjusted away.

As a result, scale indeterminacy poses a serious problem for value-added assessment. The burden of proof is on its practitioners to show that ratings assigned to educators do not vary when alternative (but in all essential respects, equivalent) scales are used. That burden of proof has not been met. Further research may yet find a way forward, but I am skeptical.

Meanwhile, there are other ways of using test data to assess schools and teachers. As states adopt comprehensive testing programs in response to the federal No Child Left Behind legislation, large databases will be assembled that make it feasible to compare the progress of any given student with a peer group that has a similar history of test results. Teachers will be able to answer the following kinds of questions: Did my students who started at the upper end of the distribution make progress equal to the statewide median gain for such students? If not, was their progress at least equal to the gain of 40 percent of the comparison group? Thirty percent? Answers to such questions should tell teachers a good deal about where they have succeeded. Moreover, because the information needed to answer these questions is limited to students’ rankings, whether the test scale has the equal-interval property is irrelevant.

California schooling

Michael Kirst (“Swing State,” Feature, Summer 2002) doubts that California has the will to stay the course and improve on the essential components of a comprehensive accountability system.

Kirst is too pessimistic. In 1995, California initiated a process that led to comprehensive K–12 academic standards, a statewide norm-referenced test, and the development of a criterion-referenced exam aligned to the standards. Three years ago the state implemented a strong accountability system. Through two education-oriented governors (Republican Pete Wilson and Democrat Gray Davis) and much hand-wringing by the education establishment, the essential features
Kirst asserts that the low-performing schools initiative disaggregates test scores “by race and ethnicity, but schools need only raise their overall scores to avoid the intervention program.” Wrong. California law is very clear. In order for schools to receive financial awards for high performance or to avoid sanctions, they must demonstrate progress for all socioeconomic, ethnic, and racial categories and for students learning English. The purpose of the law is to make sure schools do not neglect the academic needs of groups of students who traditionally perform poorly.

Regarding low-performing schools, Kirst argues, “It is less clear how a [low-performing] school can get itself out of the doghouse.” No, it is very clear: such schools must meet the annual growth targets called for under the Academic Performance Index. In the first cohort of more than 400 low-performing schools, more than 100 failed to meet these growth targets and now are (rightfully, in my opinion) the target of possible restructuring. The remaining low-performing schools are not “in the doghouse.”

California certainly has a long way to go, especially for its low-performing schools and low-performing students. Refining its testing instruments and continuing to expand its commitment to professional development are, in my view, particularly high priorities. Contrary to Kirst’s suggestion, there is no turning back on our accountability system. The public and its elected representatives will not stand for it.

Michael Kirst responds: Since the state’s nonpartisan legislative analyst now forecasts a budget deficit of at least $10 billion for 2003, it is unlikely that the impressive spending for education accountability of the past two years will continue. Implementing accountability requires incentives (as well as sanctions) plus training for staff and curricular materials for schools. The huge budget deficit will make paying for this very difficult.

Gary Hart
former California Secretary of Education
Sacramento, Calif.