Assessment tasks should be redesigned to more closely resemble real learning tasks.

two purposes. Instead, efforts to pursue an agenda like Glaser's, to develop formal measures of integrated understandings, and to train teachers in informal methods will have greater integrity if such procedures are not turned into accountability devices. While the substance of accountability tests must be improved by keeping conceptions of real learning in mind, their susceptibility to distortion should not be allowed to contaminate classroom uses of student data.

Safeguarding Accountability Assessment

To restore the credibility of accountability tests, we must remove both the incentives and the means to distort scores. For example, rewarding high test scores with bonuses to schools or merit pay to teachers is a clear invitation to teach to the test. Because teaching to the test can raise scores more dramatically than can instruction designed to improve achievement (Shepard 1988), the incentive system could reward the worst practices.

We must institute procedural safeguards to signify that scores will not be used punitively and thereby protect the meaning of the test data. For example, the use of scientific sampling to report for an entire state reduces the threat to individual classrooms, because classroom results are not reported. Similarly, a fall testing date removes the ownership for scores from the teachers who supervise test administration, thereby reducing the incentive to redirect instruction to the test. When sampling is used, policymakers relinquish the ability to rank every teacher in the state by test scores but gain believable data. If school and district results are announced to the media, data should also be reported on the wealth of each community and the meaning of the test data. For example, the use of scientific sampling to rate the performance of students statewide; they served as resources later, when the protocols were made available to teachers for classroom use. By carefully delineating levels of performance and providing adequate training, we achieved high inter-rater reliability, giving Connecticut citizens confidence that different teachers would rate a student's performance the same way.

The Connecticut Mastery Tests include performance testing to assess the state's approximately 100,000 students at grades 4, 6, and 8. Students in all three grades produce a direct writing sample. As part of the language arts tests, students take notes and use them to answer listening comprehension questions in response to taped-recorded messages. On one portion of the mathematics test, all 8th graders are required to use calculators, allowing realistic problem solving that would otherwise be too time-consuming to include on a statewide test.

Currently under development is a performance assessment of Connecticut's Common Core of Learning that portrays a standard for an educated high school graduate. Scheduled for implementation in science and mathematics in a sample of high schools during 1990-91, these assessments will focus on the integration of knowledge, skills, and attitudes and will employ principles of active and collaborative learning. They may be of short or extended duration and will include the development of portfolios, simulations, extended projects, and exhibitions. A key element of the entire endeavor will be the assessment of student attitudes, attributes, and interpersonal skills in authentic contexts.

We have found that the inclusion of nontraditional and real-world performance tasks on statewide tests often increases the amount of instructional time teachers and students spend in practicing and perfecting these and related tasks. This focus on instructional time is paramount if youngsters are to gain the flexibility and adaptability necessary for today's complex world.

Joan Boykoff Baron is Education Consultant, Connecticut Department of Education, P.O. Box 2219, Room 340, Hartford, CT 06145.