

Creating Tests Worth Taking



Grant Wiggins

the tasks (like the house) fit seamlessly with both the environment and the client's aspirations, the better the design and the result.

In this article I offer some proven design tips, tools, and criteria for fashioning assessment tasks that are more enticing, feasible, and defensible — tests worth taking.

Questions and Criteria

Designers of performance assessments should use the following key questions as a tool to guide the design process:

- What kinds of essential tasks, achievements, habits of mind, or other valued "mastery" are falling through the cracks of conventional tests?
- What are the core performances, roles, or situations that all students should encounter and be expected to master?
- What are the most salient and insightful discriminators in judging actual performances?
- What does genuine mastery of each proposed assessment task look like? Do we have credible and appropriate exemplars to anchor our scoring system? Have we justified standards so they are more than local norms?
- Are the test's necessary constraints — imposed on help available from others, access to resources, time to revise, test secrecy, prior knowledge of standards — authentic?
- Do our assessment tasks have sufficient depth and breadth to allow valid generalizations about overall student competence?
- Have we ensured that the test will not be corrupted by well-intentioned judges of student work?

The Director of Research at CLASS provides questions, criteria, and suggestions for test designers who want to engage students as well as evaluate their performance.

GRANT WIGGINS

Should a test be enticing? I think so. And should tests more often be authentic simulations of how knowledge is tested in adult work and civic settings? Many of us believe so. "Performance assessment" calls upon test makers to be creative designers then, not just technicians.

In performance assessment the design issues resemble those facing the architect. There is ground to be covered (the syllabus), there are the

logistics of making the design fit the site (making large-scale assessment work in the school), and there are building codes (psychometric norms) and town elders (school board members and district testing directors) to worry about. But designers have typically avoided another, more basic obligation: the need to serve the users — in this case, students and teachers. The clients must "own" the design; form must follow function. The more

- Who are the audiences for assessment information, and how should assessment be designed, conducted, and reported to accommodate the needs of each audience? When are audit-tests appropriate and inappropriate?

These questions can be summarized and reframed to produce eight basic design criteria:

1. Assessment tasks should be, whenever possible, authentic and meaningful — worth mastering.
2. The set of tasks should be a valid sample from which apt generalizations about overall performance of complex capacities can be made.
3. The scoring criteria should be authentic, with points awarded or taken off for essential successes and errors, not for what is easy to count or observe.
4. The performance standards that anchor the scoring should be genuine benchmarks, not arbitrary cut scores or provincial school norms.
5. The context of the problems should be rich, realistic, and enticing — with the inevitable constraints on access to time, resources, and advance knowledge of the tasks and standards appropriately minimized.
6. The tasks should be validated.
7. The scoring should be feasible and reliable.
8. Assessment results should be reported and used so that *all* customers for the data are satisfied.

The suggestions and observations that follow offer further assistance to would-be designers.

Choosing What to Test

Choose exit outcomes or areas of the curriculum that now tend to fall through the cracks of conventional testing. Typical tests, even demanding ones, tend to overassess student "knowledge" and underassess

student "know-how with knowledge" — that is, intellectual performance. Auditing local tests with Bloom's taxonomy as criteria, for example, shows that synthesis is infrequently assessed at present, and is *inherently resistant* to assessment by multiple-choice tests because it requires "production of a unique communica-

requires making the essential material of a course a *necessary means* to a successful performance *end*.

Example: a 5th grade teacher assesses geography knowledge by having his students devise a complete itinerary, map, and travel packet for their favorite rock group's world tour, within certain budget, logistical, cultural, and demographic restrictions.

Another example: students are asked to design a museum exhibit around a theme studied in a history course, selecting from many real or facsimile artifacts; required to justify what is both included and excluded in the exhibit; and must seek funding from a "foundation" of teachers and peers for the exhibit.

We want to know: Can the student use knowledge and resources effectively, *to achieve a desired effect*? This is the question Bloom and his colleagues argued was at the heart of synthesis. These tasks should only be judged well done to the extent that the content is well used.

Designing the Tasks

Contextualize the task. The aim is to invent an authentic simulation, and like all simulations, case studies, or experiential exercises, the task must be rich in contextual detail. A context is rich if it supports multiple approaches, styles, and solutions and requires good judgments in achieving an effective result. One must please a real audience, make a design actually work, or achieve an aesthetic effect that causes pride or dismay in the result.

The test may be a contrivance, but it needn't *feel* like one.² Consider professional training and testing. Doctors and pilots in training confront situations that replicate the challenges to be later faced. Business and law students learn by the case method, fully immersed in the facts of real past

Typical tests, even demanding ones, tend to overassess student "knowledge" and underassess student "know-how with knowledge."

tion" that bears the stamp of the student.¹

Faculties should also consider their institutional "customers." What kinds of tasks must our *former* students master? Here, for example, is a question from a freshman final exam in European history at a prestigious college; it suggests how even our better students are often ill-prepared for real intellectual tasks:

Imagine yourself Karl Marx, living half a century later. Write a brief evaluation of the programs of the Fabian socialists and the American reformers such as T. Roosevelt to present to the Socialist International.

Think of the knowledge to be tested as a tool for fashioning a performance or product. Successful task design

All real-world performers know the target and the standards, not just their task in advance.

cases. A context is realistic to the extent that we so accept the premises, constraints, and "feel" of the challenge that our desire to master it makes us lose sight of the extrinsic factors and motives at stake — namely that someone is evaluating us. In just this way, for example, putting out a school newspaper for a journalism course doesn't feel contrived.

Here's an example of how a teacher's attempt to design a performance task evolved as a concern for context was introduced. The original task, in a global studies course, required students to design a trip to China or Japan. But what kind of trip? For what customers? With what constraints of budget or time? The teacher then refined the task so that each student had a \$10,000 budget for designing a month-long, cultural-exchange trip for students their age. Still, the purpose is too abstract. What must the tour designers accomplish? Are they trying to design a tour in the abstract or really attract tour-takers? The students were finally charged to be travel agents who develop an extensive brochure and research the cost and logistical information using a computer reservations system.

There is no such thing as perfor-

mance-in-general. To understand what *kind* and *precision* of answer fits the problem at hand, the student needs contextual detail: it clarifies the desired result, hence the criteria and standards. Too many measurement tasks have an acceptable margin of error that is arbitrary. Are we measuring body temperature or roasts in the oven? It matters. The task's standard of performance (desired precision or quality of product) should be apparent. In fact, an important oversight by the global studies teacher was her failure to give the students model tour brochures.³

Aim to design "meaningful" tasks — not the same as "immediately relevant or practical" tasks. An assessment task will be meaningful to the extent that it provokes thought and thus engages the student's interest. But a task can be engaging without being of apparent, immediate usefulness. Whether it be mysteries, debates, mock trials, putting on plays — or, for that matter, Nintendo — students clearly respond to "irrelevant" but real challenges. What do such tasks have in common? Designers need to conduct better empirical studies to discover the tasks that tap those twin intellectual needs: *our urge for efficacy and our need for meaningful connections.*

This caution about meaning vs. relevance is particularly warranted to avoid turning important theoretical problems into crude utilitarian ones. Many genuine problems do not have obvious practical value, but they nonetheless evoke interest and provide insight into student abilities. Consider two such problems, one in geometry and one in history/English:

We all know the Pythagorean theorem: $A^2 + B^2 = C^2$; but does it have to be a square that we draw on each leg? Suppose we drew the

same shape on each leg; would the areas on A and B add up to the area on C? Find other shapes that make the equation work, too, and try to derive a more general formula of the theorem.⁴

You and your colleagues (groups of 3 or 4) have been asked to submit a proposal to write a U.S. history textbook for middle school students. The publishers demand two things: that the book hit the most important things, and that it be interesting to students. Because of your expertise in 18th-century American history, you will provide them a draft chapter on the 18th century, up to but not including the Revolution, and "field tested" on some middle school students. They also ask that you fill in an "importance" chart with your response to these questions: (1) Which event, person, or idea is most important in this time period, and why? (2) Which of three sources of history — ideas, people, events — is most important? You will be expected to justify your choices of "most important" and to demonstrate that the target population will likely be interested in your book.

Design performances, not drills. A test of many items (a drill) is not a test of knowledge in use. "Performance" is not just doing simplistic tasks that cue us for the desired bit of knowledge. It entails "putting it all together" with good judgment; good judgment cannot be tested through isolated, pat drills. As one teacher put it to me a few years ago: "The trouble with kids today is that they don't know what to do when they don't know what to do." She is right — and a prime reason is that tests rarely put students in an authentic performance situation, where thinking, not just an obvious bit of knowledge, is required.

The designer's aim, then, is to avoid inventing a new round of (this time,

hands-on) isolated items. Rather, we should consider the difference between drilled ability vs. performance ability and ask: *What is the equivalent of the game or recital in each subject matter?* What does the "doing" of mathematics, history, science, art, language use, and so forth, look and feel like in context? What are the projects and other kinds of synthesizing tasks performed all the time by professionals, consumers, or citizens that can be adapted to school use?

Such tasks are always "higher-order," and we would do well to use Lauren Resnick's criteria in our search for better-designed assessments.

Higher-order thinking

- is *nonalgorithmic* — that is, the path of action is not fully specified in advance;
- is *complex*, with the total path not visible from any single vantage point;
- often yields *multiple solutions*, each with costs and benefits;
- involves *nuanced judgment* and interpretation;

- involves the *application of multiple criteria*, which sometimes conflict with one another;
- often involves *uncertainty*, because not everything that bears on the task is known;
- involves *self-regulation* of the thinking process, rather than coaching at each step;
- involves *imposing meaning*, finding structure in apparent disorder;
- is *effortful*, with considerable mental work involved.

It may help to think of this problem as the search for larger, more interrelated but complex chunks of content to build tasks around. What, for example, might be 8 to 10 important performance tasks in a subject that effectively and efficiently "map" the essential content? Vocational programs usually grapple well with this problem by casting the course objectives as a set of increasingly complex tasks to be mastered, in which the student in the last task(s) must literally put it all together, for example, build a house in carpentry.

Refine the tasks you design by building them backwards from the models and scoring criteria. A complex task is not a vague task, with the objective or specifications unknown. All real-world performers know the target and the standards, not just their task in advance; such knowledge guides their training and rehearsals. Students should never have to wonder "Is this right?" "Am I finished?" "How am I doing?" "Is this what you want?" In a "real" problem the task is ill-structured but well-defined: the goal, specifications, or desired effect is known, but it is not obvious how to meet it. Knowing the requirements of task mastery — the "specs" — means the student must be habituated by testing to think of mastery as control over the *knowable* essentials, not as calculated cramming and good guesses. This requires providing the student with scoring criteria and models of excellent performance or production as part of instruction. (Think of diving and debate.) Such practice is the norm throughout Carleton, Ontario, where students work from "exemplar booklets" to practice grading student work — in the same way now reserved for judges in our assessments.

"What does mastery at the task look like? What will we be able to properly infer from the collected student work?" These become the key questions to ask in the challenge of taking a basic idea and making a valid performance-assessment task out of it (as opposed to an instructional task). The questions properly focus on judging anticipated results and move away from design that produces merely pleasant or interesting work.

Scoring Considerations

Score what is most important for doing an effective job, not what is



Grant Wagoner

easiest to score. The scoring rubrics should represent generalizations about the traits found in an array of actual performances. But too often we resort to scoring what is easiest — or least controversial — to observe. A fine task can be rendered inauthentic by such bogus criteria.

Two key questions for setting up a scoring system therefore are: "What are the most salient characteristics of each level or quality of response?" and "What are the errors that are most *justifiable* for use in lowering a score?" Obvious successes and errors (such as those that relate to spelling or computation) are not necessarily the most accurate indicators of mastery or its absence.⁶ Too many essay scoring systems reward students for including merely *more* arguments or examples; quantity is not quality, and we teach a bad lesson by such scoring practices.

When possible, scoring criteria should rely on descriptive language, not evaluative and/or comparative language such as "excellent" or "fair." Judges should know specifically where in performance to look and what to look for. The ACTFL foreign language proficiency guidelines and the Victoria, Australia, "Literacy Profiles" are perhaps the best examples available of such empirically grounded criteria.⁷ Teachers may also want to have students analyze a task and help devise the scoring system. This builds ownership of the evaluation, makes it clear that judgments need not be arbitrary, and makes it possible to hold students to higher standards because criteria are clear and reasonable.

"Benchmark" the standards for performance to ensure that your scoring standards are wisely chosen and suited to wider-world or next-level demands. Standard-setting for

performance involves selecting exemplary samples of performance or production. The challenge is to avoid using local age-grade norms; the solution is to equate our exit-level standards to wider-world entry-level standards at desirable colleges or professions. That advice, of course, begs a more fundamental question: Whose view of excellence should count? It is at least prudent to equate local standards of scoring to some credible wider-world

or next-level standard — something routinely done in the performing arts, athletics, and vocational education.⁸ And, *every so often*, refer to next-level standards when scoring the work of younger students. (I believe Illinois was the first state to assess both 6th and 8th grade writing samples against 8th grade exemplars, for instance.)

Administering the Assessments

Since constraints always exist in testing, make them as authentic as possible. The question is not "Should there be constraints in testing?" but rather "When are constraints authentic, and when are they inauthentic?" It is often a matter of degree, but the principle needs to be maintained and defended.

Constraints facing the designer of authentic assessment tasks typically involve access or restrictions to the following resources: (1) time



Grant Wiggins

(including time to prepare, rethink, and revise), (2) reference material, (3) other people (including access to peers, experts, the test designer, and/or the judge), and (4) prior knowledge of the tasks and how they will be assessed (the issue of test security). The question then becomes: What are appropriate limits on the availability of these resources?

Traditional testing, because it involves indirect proxies for performance, requires numerous inauthentic constraints to preserve validity. The validity of most multiple-choice tests, for example, is compromised if questions are known in advance or if reference material can be consulted during the test. These habits of administration run deep; they seem obviously required. But what of the validity issues raised by denying students access to basic resources? Just what is being tested when the student cannot predict the historical periods or books

that will be assessed, or cannot consult resources while writing?

We need not keep textbooks and other materials from students if the task is genuinely authentic. For example, in many of Connecticut's performance tasks in mathematics, the key formulas are given to the student as background to the problem. And why not allow the student to bring notes to the exam? A physics teacher I know allows students to bring an index card to the exam with anything on it; the card often reveals more about the student's knowledge than the exam answers!

Too little time for performing is not always the key issue either. Is the limiting of the test to *one sitting* authentic? If writing is indeed revision, for example, why not allow writing assessment to occur over three days, with each draft graded? Many districts now do so, including Jefferson County, Kentucky, and Cherry Creek, Colorado.⁹

I am not arguing that the student should have unlimited time and access in testing.¹⁰ Let us ask: What kinds of constraints authentically simulate or replicate the constraints and opportunities facing the performer in context? What kinds of constraints tend to bring out the best in apprentice performers and producers?

Develop a written, thorough protocol that details how the task should be administered — especially so judges will know the proper limits of their interventions to student acts, comments, or questions. It is incredibly easy to invalidate performance assessment by varying the instructions, the amount of assistance provided, and the depth of responses given to inevitable student questions. Determining beforehand what is acceptable response and intervention by adults is essential; test administra-

tors must receive standard verbal responses for delicate situations, confusions, or problems that arise.

And don't forget that kids can do the darndest things with directions that aren't thought through. In a hands-on science experiment that asked whether "the sun" heated up different colored liquids at different rates, a student did not use the heat lamp provided, moved all his equipment to the window, saw it was a cloudy day, and wrote "no."

Make the tasks maximally self-sustaining and the record-keeping obligation mostly the student's. Many educators who have never seen large-scale performance assessment cannot fathom how all students can be efficiently and effectively assessed. But they assume that the teacher will have to guide activity every step of the way and record massive amounts of information simultaneously. Thoughtful preparation, designed to make the assessment self-running, frees the teacher to be a perceptive judge.

Creating a Tool Kit

Develop a districtwide "tool kit" of exemplary tasks, task templates, and design criteria for assessment tasks. Not all of us are good designers, but why should we have to be? Teachers can help their colleagues by providing a sampler of tasks and task templates. Kentucky has done this at the statewide level, providing dozens of tasks and task ideas to teachers as part of the new state performance-based assessment system. We should consider including not only current examples of model assessment tasks, but traditional performance-based challenges such as debates, treasure hunts, mysteries, design competitions, historical reenactments, science fairs, Odyssey of the Mind tasks, Scout

Merit Badges, student-run banks and stores, and so forth.

The mathematics performance assessment team of the Connecticut Department of Education has identified the following types of problems as central to its work:

- Given data on graphs, write a story that represents the data or graph.
- Given headlines or claims with background data, explain whether or not the claims are reasonable.
- Given student work containing common errors, write a response to the student.
- Given equations or number facts, write a problem that the equations or facts could solve.
- Given trends or sample data, make and justify predictions.
- Given consumer- or job-related buying, selling, or measuring situations, solve a problem.
- Given multiple or competing interpretations of given data, justify each interpretation.

Job roles provide ample opportunities for task designers to create simulations. Here are some suggestions.

- Museum curator: design museum exhibits; compete for "grant" money.
- Engineer or surveyor: bid and meet specs for largest-volume oil container; build a working roller

Assessment design is like software design: you can never accurately and fully anticipate the naive user's response.

Good teaching is inseparable from good assessing.

coaster; map or survey a region around school or in the building.

- Ad agency director: design advertising campaigns, book jackets, or blurbs for books read in class.
- Psychologist/sociologist: conduct surveys, perform statistical analyses, graph results, write newspaper articles on the meaning of results.
- Archaeologist: determine the culture or time frame of a mystery artifact or person.
- Newspaper editor and writer: write articles and editorials set in the studied historical time.
- Policy analyst: predict the future in a country being studied.
- Product designer: conduct research, design ad campaign, present proposal to panel.
- Job interviewee: present portfolio and try to get "hired" for a specific job related to skills of current course (interview conducted by other students or teacher).
- Expert witness to Congress: testify on behalf of or against advertising claims, regulation of children's TV, or current policy issue.
- Commercial designer: Propose artwork for public buildings.

Piloting and Reporting

Always pilot some or all of the test. Assessment design is like software design: one can *never* accurately and fully anticipate the naive user's response. A first design may not fit the purpose or maximally evoke the

desired knowledge; a prompt might result in irrelevant responses that are nonetheless appropriate or reasonable to the student; the logistical constraints of a context can turn out to be more daunting than anticipated; the judges may be too self-interested in the results or insufficiently trained. A pilot is the only way to find out, even if it involves only a tiny sample of performers. And the de-bugging requires a *naïve* "guinea pig" — a teacher from a different subject or a few students — if the hidden problems in the goal, directions, or procedures are to be found.

You are what you report: Make sure that your report cards, tran-

scripts, and district accountability reports relate achievement and progress to essential performance tasks and exit-level standards. Few transcripts reflect achievement in reference to outcomes. They tend to certify that tests were passed on each isolated packet of content instead of documenting what the student can do and to what level of performance. Further, a one-shot test cannot validly assess many important capacities, as the phrases "habits of mind" or "consistency of performance" suggest. Grading and reporting thus need to move toward scoring that provides a "progress" measure — that is, work scored against exit-level

Horace's School: Redesigning the American High School

*Theodore R. Sizer
Boston:
Houghton Mifflin Company, 1992*

This book is a valuable tool for a school in the midst of a major assessment/restructuring process. Presented as an extended case study centering around the fictional teacher, Horace Smith, the book follows Horace through a series of restructuring committee meetings he is chairing at Franklin High School.

The meetings accurately capture the blend of tedium and excitement characteristic of the committee process. Flowing from the discussions of what it means to be well educated and how to best provide this education are several examples of "exhibitions." These are the means whereby students demonstrate their understanding of ideas and skills underlying the school's newly devised program. The

exhibitions provide readers Sizer's best examples of performance assessments for high school students.

Interspersed with Sizer's commentary is his narrative. It is here I found him at his best. The chapter "Policy and Power" is as cogent and heartfelt a statement about reform as I have read.

The book draws from years of research and the author's work with the Coalition of Essential Schools. Sizer delineates the Coalition's "nine common principles," which recognize there is no one way for a good school to look or proceed. Likewise, there are no shortcuts in the restructuring process, especially as it seeks to challenge the underlying principles of our current schools. This book sheds light on the reform process and helps clarify the challenge. The rest is up to us.

Available from Houghton Mifflin Company, Two Park St., Boston, MA 02108, for \$19.95 (paperback).

—Reviewed by Stephen Garger, University of Portland, Portland, Oregon.

performance standards. And no worthy performance is reducible to one aggregate score. Every student ought to have the equivalent of a baseball card — many different kinds of abilities measured and a brief narrative report — if we are seriously interested in accurately documenting and improving complex performance.

Assessment's Role in School Reform

An underlying premise of this kind of assessment reveals why I believe that assessment reform is the Trojan horse of real school reform. We badly need better definitions of mastery or *understanding* to guide assessment design, curriculum design, and teacher job descriptions and performance appraisal. Circling "correct" answers to problems only test makers care about is not "knowing," nor is it the aim of teaching. Authentic tests provide a stimulating challenge instead of an onerous obligation.

Perhaps more important for school restructuring is the need to build local educator capacity and interest in quality assessment.¹¹ Genuine faculty empowerment is impossible without deep ownership of local standards and measures. Farming all these problems out to distant "experts" is a grave mistake — one rarely made in any other country. Good teaching is inseparable from good assessing. It may well be that experts can design more rigorous tests, and that correlational/predictive validities exist in standardized tests. But schooling we can be proud of and held genuinely accountable for demands more locally useful, authentic, and enticing assessments. □

¹¹Bloom, (1954), pp. 163, 175. Serious would-be test designers would do well to reread the *text* of the taxonomy, not just the

Appendix/list, as well as the follow-up handbook developed by Bloom, Madaus, and Hastings. (1981).

"The student should [have] freedom from excessive tension . . . be made to feel that the product of his efforts need not conform to the views of the instructor . . . [and] have considerable freedom of activity . . . [including] freedom to determine the materials or other elements that go into the final product." In Bloom, (1954), p. 173.

¹²See Linn, Baker, and Dunbar, (1991), for further discussion of validity design issues.

"I have watched half a dozen classes immerse themselves in this problem and beg to continue when time ran out.

¹³From Resnick (1987).

"Describing key errors and using them in the rubric is a very *different* matter than building them into test answers as "distractors."

A related issue that emerges in designing rubrics (and thus far unaddressed by measurement experts) is the difference between the degree of difficulty of the task and the desired quality of the performance — a distinction made in New York's music performance assessments.

¹⁴See Wiggins (1991).

"Yes, yes, I know the issue is *really* one of cheating. Let the teacher "sign off" on the papers, then, certify authorship, as they do in Australia and now in Vermont.

¹⁵Though many New York State tests do allow the student what amounts to unlimited time — all day — given the shortness of the test. And certifiably learning disabled students are allowed unlimited time on the SATs as well as many state achievement tests.

¹⁶See Stiggins (1991).

References and Readings

- Bloom, B. S., ed. (1954). *Taxonomy of Educational Objectives: The Classification of Educational Goals; Handbook I: The Cognitive Domain*. New York: Longman Publishers.
- Bloom, B. S., G. F. Madaus, and J. T. Hastings. (1981). *Evaluation to Improve Learning*. New York: McGraw-Hill. [A major revision of *Handbook on Formative and Summative Evaluation of Student Learning*. (1971). McGraw-Hill].
- Linn, R. L., E. L. Baker, and S. B. Dunbar. (November 1991). "Complex, Performance-Based Assessment: Expectations and Validation Criteria." *Educational Researcher* 20, 8: 15-21.
- Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools*. New York: The Free Press.
- National Council of Teachers of Mathematics. (Forthcoming). *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*. Reston, Va.: NCTM.
- Resnick, L. (1987). *Education and Learning to Think*. Washington, D.C.: National Academy Press.
- Schwartz, J. L., and K. A. Viator, eds. (1990). "The Prices of Secrecy: The Social, Intellectual, and Psychological Costs of Testing in America." A Report to the Ford Foundation, Education Technology Center, Cambridge, Mass.: Harvard Graduate School of Education.
- Stiggins, R. (March 1991). "Assessment Literacy." *Phi Delta Kappan* 72, 7: 534-539.
- Victoria Ministry of Education. (1990). *Literacy Profiles: Handbook*. Victoria, Australia. [Distributed in the U.S. by TASA, Brewster, N.Y.]
- Wiggins, G. (February 1991). "Standards, Not Standardization: Evoking Quality Student Work." *Educational Leadership* 48, 5: 18-25.
- Wiggins, G. (June 1990). "Finding Time." *Basic Education* 34, 10.
- Wiggins, G. (May 1989). "A True Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan* 70, 9: 703-713.
- Wiggins, G. (April 1989). "Teaching to the (Authentic) Test." *Educational Leadership* 46, 7: 41-47.

Copyright ©1992, Center on Learning, Assessment, and School Structure.

Grant Wiggins is the Director of Research and Programs for the Center on Learning, Assessment, and School Structure (CLASS), 39 Main St., Geneseo, NY 14454. These design ideas derive from the work of CLASS in Frederick County, Md.; South Orange-Maplewood, N.J.; Urbandale, Iowa; and with faculties in Monroe, Orange, Suffolk, Ulster, and Wayne counties, N.Y.

Copyright © 1992 by the Association for Supervision and Curriculum Development. All rights reserved.