Using Tests Productively

by E. D. Hirsch, Jr.

Are Tests Driving Our Schools?

No effort to reform and improve our schools, such as this book proposes, can ignore testing, which currently determines what states, schools, teachers, and students do. In order to receive full benefits from the federal government, schools are required to show adequate yearly progress on reading tests for all social groups. Americans should support this stipulation of the No Child Left Behind law—which has proved difficult to meet—as it is a praiseworthy inducement to fairness and accountability. Tests of academic progress are the only practical way to hold schools accountable for educating all children and are therefore essential to the twin aims of quality and fairness. Administrators who have not met the requirement often claim that lack of federal money is the reason, but we might have expected the states, with or without a federal law, to be concerned with whether children are making adequate progress each year. Finding out through standardized tests is not expensive. Making and giving tests is not a big part of a state's educational budget.

A teacher once told me that she hated standardized tests. When I asked her, "Would you think they were so bad if your students aced them?" she said, "No, then I'd love them." The remark is far from cynical. A teacher (and her students) should love to find out that they are making real progress. Many of the complaints against the No Child Left Behind law pertain to the supposedly harmful influence of intensive preparation for the standardized reading tests. Yes, the prepping (as conducted) is harmful! But a variety of other complaints against reading tests are not justified—that they distort education, or that there is an overemphasis on tests and accountability. These objections seem justified only because there is a lack of fit between the kind of education that promotes significant progress in reading and the kind of education that the schools have currently devised in their unsuccessful attempts to

raise scores on reading tests. If the schools understood how to bring all students to reading proficiency, they would certainly do so. Many of the complaints against the tests and even the need to prepare for them would then disappear.

States are now obliged to test children in reading at every grade level, starting in third grade, in order to receive NCLB benefits. Before the law was passed, they did not have admirable accountability requirements. Some states gave tests only every third or fourth year, which was problematic, since each new school year brings the child a new teacher, who needs to know where the students stand. When I taught practicing teachers in an education school in a state that required tests only every third year, they told me that few teachers wanted to teach in the grades in which the children were to be tested, because, as they rightly surmised, they would be blamed for the faults of their predecessors in prior grades. Yearly testing is essential both to keep track of each student's progress and to encourage teachers to cooperate in providing students with a coherent education in which each grade can build on the previous one.

I want to outline some facts about reading tests that are not widely known yet need to be familiar to any parent, teacher, or citizen who is interested in educational improvement. I will cut through some of the jargon surrounding testing, and I will show how we can ease standardized tests of reading to foster a rich and formative education that will meet the requirements of adequate yearly progress for all groups with flying colors.

The Flaws of State Tests

Here are fourth-grade guidelines for teaching and testing reading comprehension, as published by three representative states. (All states issue these kinds of guidelines.)

Texas

Reading/comprehension. The student comprehends selections using a variety of strategies. The student is expected to:

(A) use his/her own knowledge and experience to comprehend;
(B) establish and adjust purposes for reading such as reading to find out, to understand, to interpret, to enjoy, and to solve problems;
(C) monitor his/her own comprehension and make modifications when understanding breaks down, such as by re-reading a portion aloud, using reference aids, searching for clues, and asking questions;
(D) describe mental images that text descriptions evoke;
(E) use the text's structure or progression of ideas such as cause and effect or chronology to locate and recall information;
(F) determine a text's main (or major) ideas and how those ideas are supported with details;
(G) paraphrase and summarize text to recall, inform, and organize ideas;
(H) draw inferences such as conclusions or generalizations and support them with text evidence and experience.

New York

Students will listen, speak, read, and write for information and understanding. As listeners and readers, students will collect data, facts, and ideas; discover relationships, concepts, and generalizations; and use knowledge generated from oral, written, and electronically produced texts.

- interpret and analyze information from textbooks and nonfiction books for young adults, as well as reference materials, audio and media presentations, oral interviews, graphs, charts,

diagrams, and electronic databases intended for a general audience
- compare and synthesize information from different sources
- use a wide variety of strategies for selecting, organizing, and categorizing information
- distinguish between relevant and irrelevant information and between fact and opinion
- relate new information to prior knowledge and experience
- understand and use the text features that make information accessible and usable, such as format, sequence, level of diction, and relevance of details.

Florida

The student constructs meaning from a wide range of texts.

1. reads text and determines the main idea or essential message, identifies relevant supporting details and facts, and arranges events in chronological order.
2. identifies the author's purpose in a simple text.
3. recognizes when a text is primarily intended to persuade.
4. identifies specific personal preferences relative to fiction and nonfiction reading.
5. reads and organizes information for a variety of purposes, including making a report, conducting interviews, taking a test, and performing an authentic task.
6. recognizes the difference between fact and opinion presented in a text.

Given such guidelines, consider the Kafkaesque predicament of schools and students under the current accountability arrangements. *The tests are coming! We don't know what topics the children will be asked to read about. The tests will probe reading comprehension skills, so we must teach those skills!*

Here are some examples of questions that appear in the fourth-grade tests put out by these states. As you will see, all of them dutifully follow the criterion that the student should be able to identify the main idea.

Texas

Paragraph 4 is mainly about
- how hard it is to become a *mahout*
- how elephants move trees
- how the *mahouts* control the elephants
- how *mahouts* protect the sanctuary

New York

This article is mostly about

- how the Appalachian Trail came to exist
- when people can visit the Appalachian Trail
- who hikes the most on the Appalachian Trail
- why people work together on the Appalachian Trail

Florida

Which detail supports the author's main idea?

- Orcas ruin his fishing profits
- Killer whales travel in pods
- Orcas prey in the ocean depths
- Killer whales are intelligent

How does one prepare students to take this kind of test? The schools have decided, on the advice of experts, that they must train students in the kinds of procedures elicited by the test: Clarify what the passage means. Question the author. Find the main idea. Make inferences about the passage. Study the meanings of words. Consider which event in the narrative comes first, and which next—all the sorts of deadening exercises that Linda Perlstein observed in her report. I call the situation Kafkaesque because, like characters in Kafka, the students and teachers are doing all the things they are supposed to do, and yet after the scores are totted up, despite their dutiful efforts, they have not fulfilled the mysterious requirements demanded by the authorities: reading scores have not improved significantly. That is because the tests are not testing comprehension *strategies,* as the states and test-makers suppose. They are testing *comprehension,* which is a different matter altogether. Reading comprehension is emphatically not a universal, repeatable skill like sounding out words or throwing a ball through a hoop. General reading comprehension is a simplified conception for something complex. It is an abstraction that stands for a whole array of separate, content-constituted skills, like the skill of reading about the Appalachian Mountains or the skill of reading about the Civil War. Unlike formal decoding skill, proficiency in one reading comprehension task does not necessarily predict skill in another.[1]

The Nature of Reading Tests

As we have seen, a student's actual ability to find the main idea of a passage is not a formal ability to follow procedures that will elicit the main idea but the ability to understand what the text says. Let's look at a characteristic bit of prose from one of these reading tests.

> There is a path that starts in Maine and ends in Georgia, 2,167 miles later. This path is called the Appalachian Trail. If you want, you can walk the whole way, although only some people who try to do this actually make it, because it is so far, and they get tired. The idea for the trail came from a man named Benton MacKaye. In 1921 he wrote an article about how people needed a nearby place where they could enjoy nature and take a break from work. He thought the Appalachian Mountains would be perfect for this.

No repetitions of classroom exercises will help the test-taker who does not know what hiking is, or what low, tree-covered mountains are like (they are not like the snow-covered Himalaya-type mountains most often pictured in books), or where Maine and Georgia are. Classroom practice in strategies cannot make up for the student's lack of the background knowledge needed to understand this passage, and no instruction in strategies is required to answer the questions quickly and accurately if the student knows about hiking, the Appalachians, Maine, and Georgia. The inferences that we make when we hear or read speech are based on a situation model particular to that utterance, derived from relevant knowledge about the domain of the passage. The comprehension skills that students are supposed to learn by practicing "comprehension skills" cannot lead to high test performance, because they do not lead to actual comprehension. If students really could gain isolated "inferencing" abilities from their strategy exercises, they would not be in their Kafkaesque trap.

In principle, there is nothing wrong with the format of most state reading tests as measures of general reading ability. But because the tests have been presented as tests of formal comprehension skills, they are unwittingly unfair, because these skills are not what they are really testing. The tests favor children who happen to have domain knowledge relevant to the passages in the test. Many test-makers go to some lengths to provide the background knowledge that they believe is needed within the tests themselves. They supply pictures. They define key words before the passage starts. But these efforts do not succeed in leveling the playing field, because students who are unfamiliar with a text's subject matter are slowed down by assimilating and applying the new background information before they consciously apply strategies to try to figure out the right answer.

Conscious strategizing is slow and cumbersome. Moreover, as we have seen, it takes the mind much longer to process meanings of a text on an unfamiliar topic. Speed is slower and scores are lower for unfamiliar topics than for familiar ones. This is true for all readers.[2]

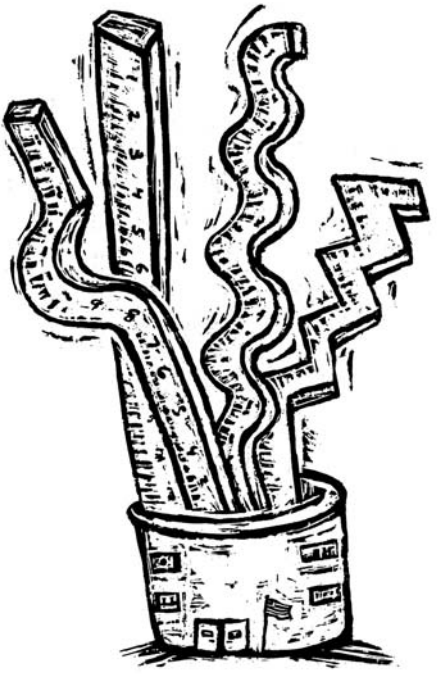
Tests are time-sensitive, as reading comprehension itself is, because slowness implies mental overload, and mental overload impairs understanding. The mental speed that is bestowed by topic familiarity is important not just for completing the test on time but also for getting the answers right. In sum, a child who already knows about the Appalachian Trail, who has heard or read about it or seen or walked on or read about similar trails, will process the passage much faster and more accurately than a child to whom such things are unfamiliar, even though the two children have identical decoding and strategizing skills. They are equally smart. They have learned equally well the lessons that the classroom has taught. Yet these two students make vastly different scores on the reading test, because one student possesses more general knowledge than the other.

There are various ways of looking at the unfairness that results when two children who have the same school-taught strategy skills receive very different scores on a reading test. Some have argued that these supposedly neutral tests are culturally biased, which is certainly true. While the test-makers attempt to be fair by making the tests knowledge-neutral, they do not succeed in this aim. Language comprehension can never be knowledge neutral. A more accurate way of perceiving the inherent unfairness of these tests is to concede that although they cannot possibly be knowledge-neutral and therefore fair to students who don't have the needed knowledge, they are perfectly appropriate as tests of reading ability. That is, their unfairness resides in the pretense that formal reading skills are being tested when in fact relevant background knowledge is being tested. Ultimately, the unfairness resides in the failure of schools to impart to all children the background knowledge they need to understand the passages on the test and similar passages in real life.

The lack of connection between the knowledge-constituted character of reading and the skills pretense of the state tests is glossed over in the technical jargon of testing, which amply confuses the general public as well as many educators. One of the main examples is the technical distinction between "norm-referenced" and "criterion-referenced" tests. The state tests from which I just extracted some main-idea items are criterion-referenced, because students in Texas, New York, and Florida are not to be measured against each other on a percentile scale, as happens with norm-referenced tests; they are to be measured against a definite criterion, namely whether they have achieved the learning goals set forth in the state guidelines for the fourth grade. They either do or don't meet the specified standard. So instead of being scored in percentiles, they are scored according to a standard of acceptability or nonacceptability, such as "proficient" and "below proficient." The point that lies between "proficient" and "below proficient" is labeled the "cut score."

This characteristic of criterion-referenced tests supposedly makes them fair and educationally productive, since everyone knows the criterion and in theory can study for it. The schools are supposed to teach to the criterion and the tests to test it.

But recall what the state criteria are: "draw inferences such as conclusions or generalizations," "determine a text's main (or major) ideas." On these empty standards, any reading test, past or present (including any norm-referenced test, as I shall show in a moment), could be considered a criterion-referenced test. Any passage on the test and its accompanying questions would fulfill the criterion as long as they include items asking the student to find the main idea and so on, *and all reading tests do this.* Thus, in the case of reading, the phrase "criterion-referenced," suggesting that the state tests are somehow based on the curriculum in a meaningful way, is very misleading. Texas reading tests would automatically meet Michigan criteria and vice versa; Florida reading tests could be used in New York and automatically meet its criteria. Why not? The state standards for reading comprehension describe empty processes. These abstract, knowledge-evasive criteria do not reflect the knowledge-based character of reading comprehension. Even if these tests were valid and reliable (an issue somewhat in doubt), they would still be inadequate when conceived as criterion-referenced tests that could productively guide schooling.[3]



Given the unfair variations in scoring from one state to another as well as the variations in the quality of state reading tests, it is worth taking a look at the makeup of well-established norm-referenced tests of reading, such as the California Test of Basic Skills (CTBS), the Stanford 9, the Gates-MacGinitie Reading Tests, the Iowa Test of Basic Skills (ITBS), and others.[4] These tests can be (and have been) used as criterion-referenced tests of reading comprehension simply by using students' actual scores instead of their percentile standings. Unlike the state tests, whose comparability with other tests is largely unknown, these standardized tests have been refined technically for many years and yield similar results when compared with one another. In reading comprehension they exhibit a correlation coefficient of around 0.8, and the correlations are even higher in the middle range of scores comprising 75 percent of students—around 0.9.[5] These high intertest correlations show that they are highly reliable—that is, they yield similar results with similar populations of test-takers. They have also been determined to be valid as measures of reading ability. (Validity means they really do measure what they claim to measure, and reliability means the scores are consistently similar with similar populations.) Their consistency has been measured carefully and at great expense. Their accuracy in measuring real-world abilities has been determined from their high correlation with a number of real-world competencies. For example, test scores in early grades predict scores in later years. Scores predict school grades. Scores predict job performance and income.[6]

If we compare the reading comprehension sections of these nationwide tests with the reading comprehension sections of the less well calibrated state tests, we find that though the tests are called by different names, they are structurally the same. The individual student's scores on any state test could be reported in percentiles, just as her scores on any percentile-reported test could be reported as being above or below a determined cut score. None of this reporting activity would change the actual score or the underlying nature of the test.

What are these tests like? Let's look at the fourth-grade reading comprehension section of one of these standard tests—the ITBS. It contains nine short passages of different genres: fiction about a bird, a biography, some lyric poetry, fiction about sports, exposition about another country, fiction about a TV program, exposition about the habits of an animal, exposition about the lives of Native Americans, exposition about a religious sect. The prose passages are short—150 to 290 words—and each is followed by around four multiple-choice questions. Now let's look at the reading comprehension component of a highly regarded state test, the FCAT of Florida. The test items have very similar characteristics—ten short, heterogeneous passages, each followed by multiple-choice

questions similar to those on the ITBS. The ITBS probably has some technical advantages, owing to its continued refinement over many years, but it is very likely that the FCAT and the ITBS will yield similar results in reading comprehension. Structurally, they are the same.

Every highly valid and reliable reading test contains several different passages sampling several knowledge areas and kinds of writing. That fact in itself gives away the knowledge-based character of reading, since if reading comprehension were a set of all-purpose formal strategies, a single passage would test reading skill perfectly well. But because general reading skill requires broad general knowledge, a valid test must sample several genres and areas of knowledge. Because of this sampling requirement, there will be no structural difference between well-designed state or national tests of reading comprehension. In short, there is no inherent difference between criterion-referenced and norm-referenced reading tests. Standardized tests like ITBS are inherently criterion-referenced, if we regard the criterion as "general reading comprehension ability." The several domains on a valid reading test are chosen not because they directly reflect what is taught in school but because they reflect an ability to read passages from an unpredictable diversity of domains. In order to read a wide array of passages in different domains, a person must have a wide array of knowledge. This is a key point, and it is currently missed in conceptualizing these tests and the instruction to prepare for them.[7]

What Kinds of Tests Will Enhance Education?

How can we calm the frantic and ineffectual test preparations of the schools and enable them to meet the adequate-yearly-progress requirement much more readily? Students and teachers cannot directly prepare for a reading test. No one should be able to predict the subject matter of the passages on such a test and specifically learn about it. That would be cheating. It would defeat the test's purpose, which is to discover how well the test-taker can be expected to read an unpredictable array of texts in and out of school. The essence of such a test is its unpredictability. But if you cannot predict a valid reading test, how can you prepare for it? You can't, and therefore you shouldn't try. You should prepare for a reading test *indirectly*, by becoming a good reader of a broad range of texts—an ability that requires broad general knowledge.

The standard reading comprehension tests, then, though adequate as reading tests, have severe shortcomings when used to measure yearly student progress in the early grades. Their two most damaging flaws are, first, that they do not positively influence instruction, since they are unrelated to any content curriculum, and second, that they cannot accurately measure yearly progress. Standard reading tests are not appropriate for

fine-scale diagnoses of the precise areas of a student's deficiencies, nor are they reliable guides to the curriculum that students should be receiving to improve their background knowledge for reading. In the early grades especially, when children are making irregular, desultory progress in knowledge and vocabulary that cannot be sensitively measured by such tests, general reading tests are quite inadequate gauges.

Like all tests, a reading comprehension test is a sampling device. It doesn't test the whole range of possible knowledge domains or kinds of text. That would make it far too long. It offers a few typical samples from a few typical domains, and students' performance on these samples is taken to predict their reading comprehension over the whole universe of reading tasks that confront the general reader. The best of the tests do a very good job of making that prediction. Although imparting the general knowledge needed for general reading ability is a multiyear project, covering at least the first six years of schooling and beyond, real progress in reading comprehension can occur in the early grades without sampling that knowledge on a reading test. If a student has just learned about the Civil War, he may not make a noticeably better grade on a short reading test that samples domains far removed from that subject. But he will nonetheless be able to read passages about Grant and Lee and Lincoln with more comprehension than he did before, even if the test does not measure that progress. He will also be able to read about events related to war and history with greater comprehension. He will know what a regiment is and what the word *bloodshed* means, though these are not on the test. He may have learned more about some of the words on the test and still not be able to answer correctly, because some of his gradual gains in word understanding, a slow, subliminal process requiring many exposures to a word, do not reach the measurement threshold of the test.

Let me quickly say that this is not an attack on these tests, and especially not on the best of them, like the ITBS. In fact, it would be sound policy to use these more established and reliable tests to measure reading ability instead of those currently made and used by the states. My real point is more radical, and, I hope, more interesting than simply criticizing tests, which are inherently necessary in education. If schools wish to meet the adequate-yearly-progress requirement, they should *systematically teach and then test for the general knowledge that leads to proficient reading comprehension.* The monitors of NCLB compliance should recognize that adequate yearly progress in early reading is in fact occurring if students show that they are not only decoding well but also gaining general knowledge, as demonstrated on curriculum-based tests of specific knowledge rather than simply on reading tests. Behind the current conception of reading, measurable, linear progress seems to be assumed. That is a reasonably correct model for the repeated, mechanical aspects of read-

ing, such as decoding. (One of the best measures of decoding skill is the ability to sound out combinations of letters that don't have any meaning at all.) But adequate yearly progress in reading comprehension cannot be accurately measured in early grades by these current tests, because much relevant learning is still latent, and the tests do not necessarily sample the knowledge areas in which progress in comprehension may have occurred.

This analysis suggests that there are far better indications of adequate yearly progress in early schooling than general reading tests that have no direct connection with the content of the school curriculum. In the early grades especially, reading tests cannot be highly sensitive measures of adequate progress in school. They are not designed to measure progress in schooling; they are designed to measure general reading ability from a sampling of subject matter that may not correspond directly to the schooling that has been provided during the year. There is a lack of fit between what needs to be taught and what is being measured.

The key to an improved test policy is to continue to use tests like ITBS and CTBS at the end of the year, as partial indicators of progress in general reading ability, especially in decoding, but to supplement them with *curriculum-based tests* that determine how well students have learned the well-defined content of the year's curriculum in all objects. We need this second kind of test to measure adequate yearly progress accurately and sensitively. Such tests would be truly criterion-based, in fact as well as in name. And they would have a powerfully beneficial effect on reading ability and on education in general. This mode of testing will encourage students and schools to learn the words and things that over time lead to reading proficiency, and at the same time it will ensure that students get proper credit for what they have actually learned and the progress they have actually made. Such knowledge-based tests are also needed to encourage and ensure long-range improvement in reading comprehension, which, as we have shown, is a skill that depends on students acquiring a wide range of general knowledge.

These content tests should be specifically tied to the knowledge goals of a sound education in literature, science, history, and the arts, for these are the large domains that constitute the background knowledge required for reading comprehension. Ideally, the curricula to which the tests are tied should be focused on the knowledge that is most important and enabling from the standpoint of later learning and reading ability. It takes several years of systematic, cumulative learning before schoolchildren can gain the general knowledge and conceptual fluency they need to be good readers. To those who might object that I am recommending more rather than less testing, I reply that content testing leads to engaging, productive, and interesting teaching, whereas drill-and-kill process testing does not.

John Bishop, of Cornell University, has shown that educational systems which require definite content standards *and* use curriculum-based content tests to determine whether the curriculum has been learned greatly improve achievement for all students, including those from less advantaged backgrounds.[8] Additional evidence in support of curriculum-based content testing comes from the recent finding that gains in reading are directly proportional to the completeness with which a school implements a coherent, content-rich curriculum.[9] A system of specific content standards coupled with curriculum-based tests will cause achievement on non-curriculum-based tests to rise over time. It will result in higher achievement overall and a narrowing of the academic gap between rich and poor.

We should abandon the formalistic conception behind current testing policy in reading comprehension. It is a self-defeating policy based on mistaken ideas, and it should be replaced by a testing policy that encourages schools to teach the general knowledge that will lead to proficient reading comprehension. Breadth of knowledge is the single factor within human control that contributes most to academic achievement and general cognitive competence. In contradiction to the theory of social determinism, breadth of knowledge is a far greater factor in achievement than socioeconomic status. The positive correlation between achieved ability and socioeconomic status is only half the correlation between achieved ability and the possession of general information. That is to say, being "smart" is more dependent on possessing general knowledge than on family background per se.[10] This little-known and quite momentous fact means that imparting broad knowledge to all children is the single most effective way to narrow the competence gap between demographic groups through schooling. The tests we give should reflect our understanding of this truth.

E. D. Hirsch, Jr., is the author of the best-selling Cultural Literacy. *He is the founder of the Core Knowledge Foundation, to which all his proceeds from* The Knowledge Deficit *will go, and a fellow of the American Academy of Arts and Sciences.*

Notes

1. Donna R. Recht and Lauren Leslie, "Effect of Prior Knowledge on Good and Poor Readers' Memory of Text," *Journal of Educational Psychology* 80(1) (March 1988): 16–20.

2. E. D. Hirsch, "Measuring the Communicative Effectiveness of Prose," in J. Dominic, C. Fredricksen, and M. Whiteman, eds., *Writing* (Hillsdale, N.J.: Erlbaum, 1981), 189–207. See also Recht and Leslie, "Effect of Prior Knowledge."

3. There is another source of unfairness in these state tests: the cut scores in the different states vary a good deal. This means that a student is deemed proficient in one state but would be deemed below proficient in another. I have been told by authorities in several states that reading cut scores are decided on *after* children take the tests, so the states can meet the political requirement for a reasonable number to pass and so the schools will not appear ineffectual. This variation in cut scores is definitely not the case for the congressionally mandated National Assessment of Educational Progress (NAEP) reading test. The cut score between "proficient" and "below proficient" is carefully determined and is applied without favor to children in all states. We can compare NAEP results in a state with the state's own results and thus gain a calibration tool with which we can determine just how meaningful and fair the cut scores are in the different states. This discloses a very large variation. It is hard to say which kind of unfairness uncovered here is more deplorable—the unfairness that students in one state fail while students with the same reading abilities pass in another, or the unfairness of passing a student who cannot understand much of what he reads. See Jennifer Sloan McCombs, Sheila Nataraj Kirby, Heather Barney, Hilary Darilek, and Scarlett J. Magee, "Achieving State and National Literacy Goals, a Long Uphill Road: A Report to Carnegie Corporation of New York," TR180EDU (Santa Monica: Rand Corporation, 2004).

4. Ibid. See also P. E. Peterson and F. M. Hess, "Johnny Can Read . . . in Some States," *Education Next* (Summer 2005): 52–55.

5. See, for example, J. I. Brown, V. V. Fishco, and G. Hanna, *Nelson-Denny Reading Test,* Technical Report, Forms G & H (Chicago: Riverside, 1993); W. H. MacGinitie and R. K. MacGinitie, *Gates & MacGinitie Reading Tests,* 3rd ed., Technical Report (Chicago: Riverside, 1989).

6. William R. Johnson and Derek Neal, "Basic Skills and the Black-White Earnings Gap," in Christopher Jencks and Meredith Phillips, eds., *The Black & White Test Score Gap* (Washington, D.C.: Brookings Institution, 1998), 480–497.

7. John B. Carroll, "Psychometric Approaches to the Study of Language Abilities," in C. J. Fillmore, D. Kempler, and S. Y. Wang, eds., *Individual Differences in Language Abilities and Language Behavior* (New York: Academic, 1979).

8. J. Bishop, "The Impacts of Minimum Competency Exam Graduation Requirements on High School Graduation, College Attendance, and Early Labor Market Success," *Labour Economics* 8(2) (2001): 203–222; J. Bishop and M. Bishop, "How External Exit Exams Spur Achievement," *Educational Leadership* 59(1) (2001): 58–65.

9. G. Taylor and G. Kimball, "The Impact of Core Knowledge Implementation on Student Achievement in the Oklahoma City Public Schools," Occasional Papers, Oklahoma City Public Schools, May 2000, available at <http://www.coreknowledge.org/CKproto2/about/eval/eval2_2002.htm>; "Summary of Research on the Effectiveness of Core Knowledge," available at <http://www.coreknowledge.org/CKprot02/about/eval.htm>.

10. D. Lubinski and L. G. Humphreys, "Incorporating General Intelligence into Epidemiology and the Social Sciences," 24(1) (1997): 159–202. The positive correlation between achieved ability and socioeconomic status is 0.422, whereas the correlation between achieved ability and general information is 0.811.