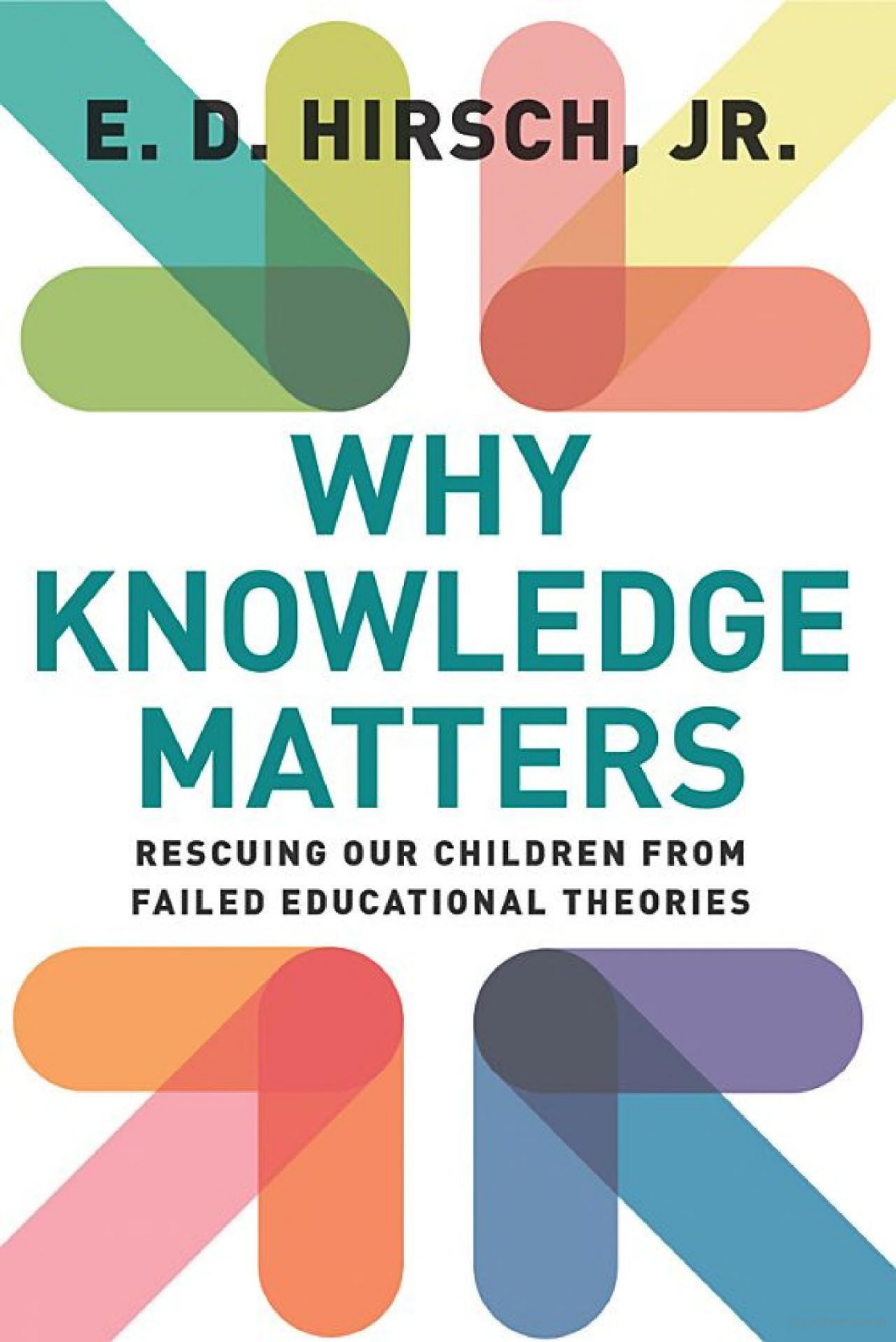
E. D. HIRSCH, JR.
WHY KNOWLEDGE MATTERS
RESCUING OUR CHILDREN FROM FAILED EDUCATIONAL THEORIES

Paperback ISBN 978-1-61250-952-5
Library Edition ISBN 978-1-61250-953-2
ePUB ISBN 978-1-61250-954-9

Library of Congress Cataloging-in-Publication Data

Names: Hirsch, E. D. (Eric Donald), 1928- author.
Title: Why knowledge matters : rescuing our children from failed educational theories / E. D. Hirsch, Jr.
Description: Cambridge, Massachusetts : Harvard Education Press, [2016]. | Includes bibliographical references and index.
Identifiers: LCCN 2016012718 | ISBN 9781612509525 (pbk.) | ISBN 9781612509532 (library edition) | ISBN 9781612509549 (ePUB)
Subjects: LCSH: Education, Elementary—Curricula. | Educational change. | Educational accountability. | Education—Standards. | Education—Aims and objectives. | Educational innovations.
Classification: LCC LB1570 .H57 2016 | DDC 372—dc23 LC record available at https://lccn.loc.gov/2016012718

Published by Harvard Education Press,
an imprint of the Harvard Education Publishing Group

Harvard Education Press
8 Story Street
Cambridge, MA 02138

Cover Design: Wilcox Design
Cover Image: iStock.com/jamesjames2541
The typefaces used in this book are Adobe Garamond Pro and Futura.

CONTENTS

PROLOGUE

The Tyranny of Three Ideas

THE FRENCH CONNECTION

These chapters are organized around six well-recognized educational frustrations in the United States: the over-testing of students, the fadeout of preschool gains, the narrowing of the elementary curriculum, the low verbal scores of high school graduates, the lack of progress in closing achievement gaps between social groups, and the tribulations of the Common Core initiative. These problems have defied solution—but not primarily from lack of will or money, or from poverty or the shortcomings of teachers. They resist solution because they cannot be solved under the reign of the faulty ideas that caused them to arise. My long-standing view is that idea change will be the most effective educational reform of all.

Those who know my past work may wryly object: *you* haven't changed *your* ideas for thirty years.[1] That's both true and false. The basic themes are largely unchanged. The hedgehog knows one big thing.[2] I am still chiefly motivated by the social injustice of our dominant theories and their unwitting destruction of the American dream. On that topic, I say with Matthew Arnold: "Charge once more then, and be dumb."[3] The reader will find the theme of equalizing opportunity a leitmotif of the book in all of its chapters.

Against the tide of sociological and genetic explanations of achievement gaps, the path-breaking work of researchers like John Guthrie, Betty Hart, and Todd Risley should have made it unnecessary to assert once more at this late date that the achievement gap is chiefly a knowledge gap and a language gap. It can be greatly ameliorated by knowledge-based schooling.

Once the centrality of knowledge (not general "skills") is fully grasped by educators and the wider public, the right to parity of knowledge among young pupils will come to be understood as a civil right. This book continues my earlier theme that only by systematically imparting to *all* children the knowledge that is commonly possessed by successful citizens can all children gain the possibility of success —"success" understood as becoming a person with autonomy, who commands respect, has a communal voice that can write and speak effectively to strangers, can earn a good living, and can contribute to the wider community.

But this book is far more than a rehash of former ideas about what is needed for equality of educational opportunity. New findings in cognitive science have helped me gain greater clarity and depth. The book has benefited from my clearer understanding that the key task facing our elementary schools is to shift our emphasis from the goal of self-realization to the goal of community—from child-centeredness to community-centeredness. No sensible person would disparage either goal. But the emphasis must shift decisively for the sake of the community *and* the individual child.

With this book, I hope to reach readers who had barely

come into the world when my *Cultural Literacy* was published in 1987. We live in an era of new possibility. We have witnessed the failures of recent educational theories, but at the same time we have also witnessed marvelous new modes of spreading knowledge—should better theories be adopted. The great physicist Max Planck, the progenitor of quantum physics, despaired of ever convincing his fellow professors to change their views. He looked to the young. He complained that professors never change their minds; they die off, and the younger generations take their places.[4] And indeed some young scholars have recently begun to invoke my name in the blogosphere as a kind of superannuated mascot. The context and the national mood have changed. Heterodox ideas that were rejected a few years ago might now be granted a new hearing after the frustrations of current reform efforts.

The most immediate impetus for this book is my discovery of shocking new evidence on these issues from France. There is a radical streak in French thinking that encourages sudden and complete national transformations—the French Revolution being only the most famous instance. For many decades the French elementary school had been the pride and the terror of the young, with every child, rich or poor, having to undergo the very same rigors under the same national curriculum. The egalitarian impulse of this uniformity was expressed early in the Revolution by Condorcet in his 1790 pamphlet *A Common Education for Children*, and re-expressed in the nineteenth century by Jules Ferry, the founder of modern education in France. In his 1883 letter to teachers, Ferry urged them to teach "that knowledge which is common to all and indispensable to all." Those sentiments were reconfirmed

in 1977 by the centrist president of France, Giscard d'Estaing, who stated, "The defining and acquiring of the very same knowledge by all French children, who from now on will all go to the same primary school, and the same middle school, will be an essential element in the unity of French society, and in the reduction of inequalities of opportunity."[5] But in 1989, the bicentenary of the French Revolution, France passed a radical new education law—the *loi Jospin*—requiring all elementary schools to cease teaching the national curriculum and begin teaching locally determined curriculums, individualized further by a special emphasis in each school, called its *projet*. This drastic change had been silently prepared for during two decades of teacher indoctrination within French education schools into American-style progressive education. The new law reflected those ideas: more attention was to be paid to the individuality of each student, to his or her native abilities, interests, and home culture. To compensate for all this novel heterogeneity, the unifying emphasis was to be on general skills such as "critical thinking" and "learning to learn." In other words, in 1989 the French decided to completely Americanize their school system overnight.[6]

The sudden organizational change introduced by the new *loi Jospin* instituted a vast natural experiment. Which mode of schooling would work better and more fairly: the community-centered and knowledge-centered mode of the past, or the child-centered and skills-centered mode of the future? The broad new law enabled the Ministry of Education to conduct longitudinal studies comparing the effects of the communal elementary curriculum before 1989 with those of the

individualized, skill-centered curriculums that followed.[7] It was a natural experiment because many key elements of French education, other than curriculum, stayed constant over time. Teacher quality stayed the same by objective measures. School buildings and budgets did not change significantly. The superb French preschools were not covered by the new law, and stayed essentially the same. The most decisive change was in the curriculum and pedagogy of the elementary school.[8]

Ministry researchers have now analyzed the results over twenty years among various demographic groups. Their data was gathered from ten-year-olds at the end of primary school. They reported an astonishingly steep decline in achievement in each demographic group—children from the homes of wealthy executives and professionals, children from the middle classes, children from various other well-defined demographic groups including the unemployed, with their ever-higher percentages of immigrants from North Africa. Each group was academically harmed by the new system, and the harm became ever greater as one went down the economic scale. The children of the unemployed declined most of all. Achievement decreased. Inequality increased dramatically.

The massive declines that occurred at the very top among children of white-collar workers and high-level professionals and executives cannot be blamed on the influx of North African immigrants, as some American experts are inclined to say.[9] Why are American education experts inclined to blame immigrants for the French decline? They know little about the details. This book contains the first extensive discussion

in English.[10] It would certainly be reasonable to blame a big influx of immigrants for a decline in the *average* of French test scores. But a fair-minded person would hardly blame the children of immigrants (who suffered most of all from the new regime) for a big decline among the children of native-born executives and professionals.

An entire educational theory has been put to the test in France, with incontrovertible results that everyone in France now calls "the crisis of the school." The American-style, individualistic theory yielded far worse results for every demographic group. As a tenacious theory holder myself, I can't blame educational experts for seeking an alternative explanation. But I'd like to believe that I'd be willing to give up my theory rather than resist such decisive evidence. Compare this French research with our own best research—for example, our longitudinal analyses conducted by the National Assessment of Educational Progress (NAEP). That research is based on a sample size of eighty-seven hundred students from an age cohort of six million. The French longitudinal studies are based on a sample size of forty-two hundred students from an age cohort of one million. By those numbers alone, even without the refinements introduced by French and American experts, the two samplings yield very similar levels of confidence.[11] I will use the details of the French results throughout this book, and devote chapter 7 to an analysis of the French experience.

THE SILVER AGE

Underlying these chapters is a historical narrative—the story of a decline in American schooling to be followed by a

renewal, if we are wise.[12] The practical policy changes that I will advocate are founded on ever-stronger scientific evidence and an ever-clearer picture of that historical narrative. It is folly to pretend that our historical mistakes are irrelevant to the problems we currently confront and the policies we need to put in place. There were past causes of our educational decline, and there are still-current reasons why we have not recovered from it. The verbal scores of our seventeen-year-olds have stayed low and largely unchanged ever since NAEP began recording them in 1971. But by that year, the decline was already in full swing. The ideas that caused the decline still remain in full force today. These historical facts, coupled with recent cognitive research, will add credibility to the view that our educational fate is largely controlled by ideas rather than by irresistible social forces.

The decline in our student test scores in the 1970s was caused by the dominance of conceptions that had begun to take over American public schools starting in the 1920s. The ideas did not complete their conquest right away. As late as the 1940s and 1950s the public education of the United States, for all its racial and social shortcomings, scored near the top among nations in both achievement and equality.[13] Moreover, as John Bishop has long pointed out, the education gap between blacks and whites had narrowed steadily until recent decades.[14] But between 1960 and 1980 American academic scores fell rapidly *at all grade levels*—more than 25 percent of a standard deviation, a big drop for large populations.[15] The verbal scores on the SAT fell 50 percent of a standard deviation! Those puzzling disasters caused the Reagan administration to convene a national commission that

produced the famous alarmist report *A Nation at Risk* (1983). When the French later adopted those same ideas they suffered a decline of similar massive proportions. Our "*Nation at Risk*" of 1983 became their "*Crise de l' école*" of 2007.

The belief that there was once a golden age of American education is scorned by educational historians. They are of course right. But they concede that there was indeed a large test score decline—over a quarter of a standard deviation—in grade school and high school test scores between 1960 and 1980.[16] As this book shows, the decline occurred at all grade levels among all demographic groups.[17] Thereafter, in our own times, test scores have remained low and stable within a tenth of a standard deviation.[18] So let's call the higher-scoring era before the decline—the 1940s and 1950s—a "silver age." The subsequent test-score decline and its causes are important to know about, acknowledge, and rectify.

The chief cause of the decline was the nationwide adoption of a set of inadequate ideas.[19] Though the ideas were partly true and beneficial, they were also partly incorrect and harmful because they neglected the communal dimension of education in favor of individualistic child-centered development. The French have now repeated our experiment in educational individualism in a more concentrated and better-documented form.

Here are the three basic ideas that depressed education in both nations:

- Early education should be appropriate to the child's age and nature, as part of a natural developmental process.
- Early education should be individualized as far as possible—to follow the learning styles and interests of each developing student.

- The unifying aim of education is to develop critical thinking and other general skills.

The new policies that I (and others) recommend are based on a different set of ideas and emphases that are more consistent with current cognitive science, developmental psychology, and social science:[20]

- Early education should be chiefly communal—focused on gaining proficiency in the language and the conventions of the public sphere.
- Every child in each locality should study basically the same early curriculum.[21]
- The unifying aim of early schooling is autonomy and equality of opportunity: to impart to every child the enabling knowledge that is possessed by the most successful adults in the wider society.

No doubt our current principles—natural development, individuality, and critical thinking—will continue to be regarded with favor by many people. The ideas are attractive. They counsel empathy with the individual child, and they claim to comport with the child's natural development. Naturalists will of course concede that communal knowledge is important, while communalists will concede that nature cannot be thwarted.[22] That agreement sounds very promising. But emphasis is critical, and foes of inequality like me caution that if an advantaged child at age seven knows certain things without harm, then it *cannot* be inherently harmful or "developmentally inappropriate" for a disadvantaged child also to know those very same things at the same age.

And the communalist will further caution that there is a big distinction between accommodating shared curriculum

topics for each child as the best schools do in the community-centered schools of Finland and Japan, and devising different curriculum topics for different children as we and the French now do in the child-centered school. Elementary school is a time for building socialization as the only means through which individuality can ultimately express itself. Children need to master the shared conventions of the standard language and of social interaction.[23] They need to learn the shared knowledge and vocabulary of the nation, the shared spelling, pronunciation, and other conventions in the public sphere of the grown-up world. Only full membership in the tribe leads to individuality, as G. H. Mead profoundly observed.[24]

Caricatures of the communal view dismiss it as "lock-step education," "indoctrination," "one size fits all," "the factory model of schooling." But I will show in chapter 1 that, paradoxically, it is the naturalistic and individualistic view that has turned schools into soulless test-prep factories, with endless practice of strategies and skills, as they desperately attempt to overcome children's lack of enabling knowledge—a lack partly induced by an individualized rather than a communal curriculum. I hope that my recommendation of a shift in emphasis from individual to community will not be misunderstood as lack of affection and solicitude for the individual child. On the contrary, our assumptions about how children learn have led to instruction that is far from child centered, and that perpetuates inequality among children from different backgrounds.

Old-timers in education reform might suppose that when I use the phrase *communal curriculum* I am implicitly promoting

the Core Knowledge Sequence for the early grades—a coherent, cumulative, and content-specific curriculum guide offered for free on the Core Knowledge website.[25] New readers need to be aware that I started the Core Knowledge Foundation back in 1986. After four years of labor and consultation the Foundation produced the Core Knowledge Sequence for preschool through grade eight in 1990. How it was created is described in the introduction to the Sequence.[26] That Sequence proposes to teach everyone the enabling knowledge (including up-to-date, multicultural knowledge) shared by the most successful adults in America today.

But the promotion of any single curriculum guide has been far from my mind, and is not a motivation for this book. Rather, this book's aim is to promote the general communal principle. The Core Knowledge Sequence has always been offered as just one exemplification of the more general idea that there exists a de facto public commons that enables our national language to be deployed effectively, and that every child in a democracy should have access to that shared, enabling knowledge and language. No matter what the home culture might be, every child deserves to become proficient in the taken-for-granted knowledge of the standard language. The main mission of the Foundation is to serve that general communal idea, which can be realized by different curricula that vary in interesting ways.[27]

I have recently begun to name that general principle "communal knowledge." Whole nations have successfully followed communal knowledge in the form of national curriculums that have a similar communal purpose. No large

nation has done so more successfully than France did from 1975 to 1985, when it had the highest achieving, most egalitarian school system of any large country in Europe. After 1989, the French in effect duplicated the American decline of the 1960s and 1970s by means of the same basic change in guiding ideas. The Americans, of course, never had a national curriculum like the French, but the schools of most American districts did in earlier days have a strong communal purposiveness.[28]

Education without an explicit communal purpose is unlikely to achieve a communal result that offers every child economic competence and entrée into the public sphere. The adoption of more communal ideas than those that now prevail in the United States and France could offer both nations a new birth of fairness and excellence.

TWO CHEERS FOR THE THREE PREVAILING IDEAS

Any idea such as developmental appropriateness and child-centeredness that keeps earning the adherence of teachers all over the world must have a strong tincture of truth. That's surely the case with two out of the three guiding ideas of current American elementary education: naturalism and individualism. But the third guiding idea, which one could call *skill-centrism*—the aim of imparting critical-thinking skills and similar general skills like problem solving—is altogether problematic.

Naturalism and individualism go together.[29] They arose from a belief that nature, as the earthly manifestation of a beneficent God, is unerring and benign. Hence the natural growth of a child is an instance of God unfolding His purposes

in the world. (The root meaning of *development* is "unfolding.") So nature cannot betray. It is the true guide that will lead to physical and spiritual health. And since each child's nature is special and different, following nature will mean adjusting education to the naturally developing interests and abilities of each child. I have adopted the phrase *providential individualism* to capture this point of view. I have found it useful in describing the widely held faith that if we let affairs take their natural course we are in the hands of a benign Providence, so all will be well, even optimal, in education. The source of this faith is the unspoken assumption that a benevolent purpose is present in Nature, and will assure a beneficent result.[30]

This naturalism plus individualism is emotionally compelling. It is reinforced by our love and solicitude for young children. It leads to empathetic teaching, since love and concern for the individual child is a more sustaining and agreeable mode of instruction than fear. Of course, naturalism and individualism have no monopoly on a loving and empathetic teaching, which is in all cases the best pedagogy for young children.

But an implication has been drawn from providential individualism that has created a serious problem for American education. Naturalism and individualism, beyond implying a loving pedagogy, have also been taken to imply—and this is a fatal weakness—a *curriculum* that arises from the child's individual abilities and temperament: "different strokes for different folks," "multiple learning styles," "multiple intelligences." American school mission statements usually proclaim that the school will provide an education

tailored to the individuality of each child.

But I will argue, with support from developmental psychology, that equating early education with the metaphor of individual "development" is misleading; that so-called "unnatural" social impositions are the most natural things in the world; that school systems with so-called "lockstep" curricula in the early grades (Finland, Japan) have very child-happy effective schools that score near the top in international studies.[31] Indeed, international studies have shown that a differentiated curriculum is harmful to achievement and equity.[32] To make the emphases and content of the child's early schooling largely dependent upon the child's uniqueness is an idea unsupported by developmental psychology.[33] The evidence for individual learning styles is weak to nonexistent.[34] And in practice the individualizing of the elementary-school classroom has led to fragmentation of the curriculum.

This fragmentation is defended and supposedly turned to benefit by a third doctrine: that the goal of education is the imparting of general skills like critical thinking, creative thinking, problem solving, and cooperative thinking. But reality has not accepted this hopeful idea about skills, and recent cognitive science has been fatal to it.[35]

Educational individualism has always required the general-skills idea. To make thinking skill the ultimate goal renders irrelevant the fragmenting of school topics that must occur when the teacher is urged to tailor the curriculum to the uniqueness of each child.[36] Current thinking holds that the fragmenting of the early curriculum will work out in the end, because the goal is not chiefly to impart the specific content

of the curriculum but rather to train the mind to critical thinking and problem solving for any content. This connection between the general-skills idea and individualism in the curriculum was the subject of a 1910 book by John Dewey called *How We Think*. He says this in his preface: "Our schools are troubled with a multiplication of studies, each in turn having its own multiplication of materials and principles. Our teachers find their tasks made heavier in that they have come to deal with pupils individually and not merely in mass. Unless these steps in advance are to end in distraction, some clue of unity, some principle that makes for simplification, must be found. This book represents the conviction that the needed steadying and centralizing factor is found in adopting as the end of endeavor that attitude of mind, that habit of thought, which we call scientific."[37]

This statement has seismic importance for understanding recent American educational history. By no means should Dewey be scapegoated for articulating this central idea in 1910. He is stating a practical necessity: if the content of the curriculum is to be scattered and diversified by "dealing with pupils individually and not in mass," then some further principle is needed to guide instruction and lend unity to the experience of the individual student. That can only be accomplished, Dewey says, by making critical thinking rather than mere facts the proper goal of child-centered education. Dewey is right about the structure of the difficulty, and he has also identified what may be the most recalcitrant political problem in American education—that few dare challenge our emphasis on individualism.

The proposal that critical thinking is an aim that unifies

fragmented and individualized schooling made sense in Dewey's era, when scientists had incorrect ideas about skill development. But research on thinking skills is now a well-developed field, and its findings are fatal to this crucial refuge of current educational theory. Here's a brief summary of findings from a recent book on the subject, *The Cambridge Handbook of Expertise and Expert Performance* (2006): "Research clearly rejects the classical views on human cognition in which general abilities such as learning, reasoning, problem solving, and concept formation correspond to capacities and abilities that can be studied independently of the content domains."[38]

Modern cognitive psychology holds that the skills that are to be imparted to a child by the school are intrinsically tied to particular content domains. This is called the *domain specificity* of skills. Thinking skills cannot readily be separated from one subject matter and applied to other subject matters. The domain specificity of skills is one of the firmest and most important determinations of current cognitive science. The Cambridge compendium from which the passage is taken is *not* called *A Handbook of Skills*, which could imply all-purpose skills. It's called a *Handbook of Expertise*, implying that the basis of skills is specific domain knowledge. Think of how significantly our view of schooling might change if suddenly policy makers, instead of using the term *skill*, had to use the more accurate, knowledge-drenched term *expertise*.[39]

Dewey's worry was well founded. The principle of unity was devised to support child-centered education and keep it from ending in fragmentation. Yet that single, overarching skill doesn't exist. Believing in that mirage has actually

resulted in the "distraction" Dewey feared. It has induced an ever-more-desperate effort to gain nonexistent skills through soul-deadening drills. A benign child-centeredness coupled with a faulty theory about general skills has led us to a child purgatory of skill drills. These have produced neither good skills nor good scores on the ever-looming tests.[40] Those distracting tests will be the subject of my first chapter.

CHAPTER ONE

The Invalid Testing of Students

THE STORY OF ASTERISKS

This chapter will be critical of our current reading tests. The public agrees: a furious outbreak of antitesting sentiment has broken out over the nation, with parents and students boycotting required tests. This kind of activism seems misguided. Tests have always been necessary in education. A better solution is to make tests fewer and better. There's just one way to do that—to base them on well-defined, knowledge-based curriculums. There is no other way of making tests fair and productive. It's also the only way to make schools excellent and fair. That will be the long and short of this chapter.

If one had to choose a single measure of the academic quality of a school system, the average reading score of its graduating seventeen-year-olds would serve. Verbal scores at that age predict students' college and career readiness and their later economic success.[1] A technically valid reading test that accurately predicts a student's ability to comprehend diverse texts will also, self-evidently, predict that student's ability to comprehend both oral and written language. A reading test is a test of general knowledge and vocabulary; it gauges a student's ability to operate effectively within the public sphere. Since a valid reading test probes a student's

degree of initiation into the public sphere—a fundamental aim of education—any policy that lowers or neglects to improve test scores in reading is a failed educational policy.

In math, in contrast to reading, American scores for seventeen-year-olds have been stable for many years. While it's disappointing that math scores at age seventeen haven't improved markedly, at least they haven't gone down, as reading has. For, in 2012, a decade after the test-intensive No Child Left Behind Act (NCLB) went into effect, the reading scores of seventeen-year-olds came in significantly lower than they had been in 1988 before NCLB was enacted. Figure 1.1 is taken from the 2012 long-term trend report of the National Assessment of Educational Progress (NAEP)—the latest nationwide analysis we have for long-term reading trends among seventeen-year-olds. It shows a statistically significant drop of three points in the reading abilities of seventeen-year-olds, as compared with scores in 1988, 1990, and 1992.

Notice the asterisks in the chart—a valuable feature of the long-term reports. They denote a statistically significant difference (lower or higher) between earlier results and those of 2012. Whether a prior score gets an asterisk is the result of data analysis that goes deeper than the surface averages.

This 2012 result for seventeen-year-olds seems particularly anomalous, because that same cohort four years earlier, at age thirteen, had attained the topmost score that NAEP had ever recorded—higher than any prior cohort except 1992, with which it was tied. We see such anomalies across the years. Thirteen-year-olds have made steady progress in reading, but the scores of seventeen-year-olds have remained

flat. No college, or employer, or nation much cares how well students did at age thirteen if by age seventeen their verbal abilities are worse than they were before the new reforms were instituted.

Could our recent high-stakes testing regimens have contributed to this disappointment? Students who were seven in 2002 would have been old NCLB hands in 2012. They would have experienced almost ten years of high-stakes testing in reading under NCLB. Those testing regimens will have deeply affected what schools and teachers taught them in their early grades. Yet all that intensive test prep did not, in the end, help their mature reading. Since their age group scored significantly better before the reign of NCLB, we might ask: Could the well-meant tests have actually promoted long-range educational harm?

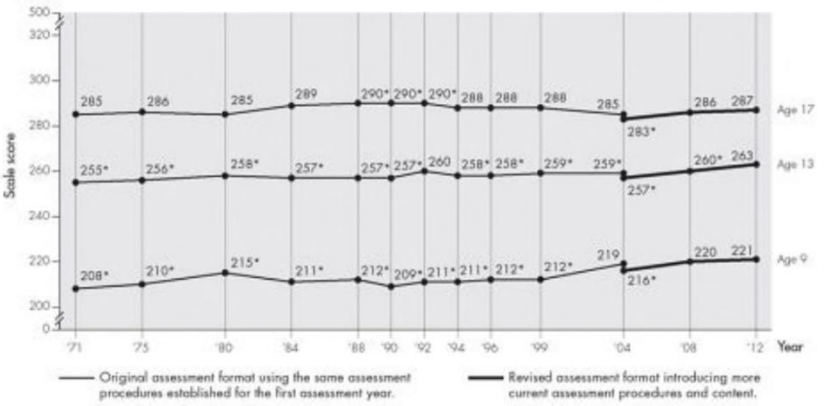

FIGURE 1.1 Trend in National Assessment of Education Process reading average scores for nine-, thirteen-, and seventeen-year-old students
Source: US Department of Education, Institute of Education Sciences, National Center for Education Statistics, NAEP, various years, 1971–2012 Long Term Trend Reading and Mathematics Assessments
*Significantly different ($p < .05$) from 2012

That's a troubling thought to me, since the testing regimens have clearly helped improve the mechanics of early reading (that is, the ability to sound out fast and effortlessly)—an important gain. Younger students can now decode texts with more fluency and accuracy than they did before NCLB. Because of that the reading gaps between ethnic and racial groups have narrowed in those years, when fluency is an important part of the score. But, alas, those equity gaps begin widening again between ages thirteen and seventeen, when knowledge and vocabulary are decisive.

One might first be inclined to blame the high schools for undoing what the grade schools had achieved in reading. But as a specialist in reading comprehension I know that such an accusation is incorrect. The vocabulary size of seventeen-

year-olds is not determined at age seventeen, or sixteen, or fifteen. It's a plant of slow growth that is determined by the knowledge that has been gradually acquired from a child's overall experience from birth to age seventeen. Early schooling can play a major role in vocabulary growth, especially for disadvantaged children, and some forms of schooling do much better than others. While current schooling in early grades has improved the sounding-out aspects of reading, we can infer from the NAEP results that it has depressed the vocabulary sizes of seventeen-year-olds. My conclusion, which sets the stage for this chapter, is that too much time is now being spent on test preparation in early grades, and too little time is being spent on gaining the wide knowledge required for a broad vocabulary.

To summarize the story of the asterisks: high-stakes tests, upon which the lives of students, teachers, principals, and superintendents now anxiously depend, became by slow degrees a feature of American education in the decades after *A Nation at Risk* of 1983. That report showed that American public schooling had declined in quality during the 1960s and 1970s. After the governors' education summits of 1989 and 1996, all but one state (Iowa) developed grade-by-grade reading standards on which schooling in the state would be based. Then the No Child Left Behind Act, which was signed into law in 2002, mandated standards and yearly high-stakes testing in reading and math based on the state standards.

I mentioned one consequence of these reforms: an improvement in the teaching of decoding (i.e., turning written marks into sounds and words). That in turn should also have produced an improvement in reading

comprehension, and it did so for undemanding texts on everyday topics in the early grades. But it has not produced improvement in the reading comprehension of more demanding texts on more demanding topics. I was mentally prepared to look for this anomaly because a number of schools told me, after high-stakes reading tests came into effect under NCLB, that they could no longer teach history and the arts. They now were being made to teach "reading" instead, with a strong emphasis on test preparation. Their higher-ups were under the impression that intense classes devoted to "making inferences" and "finding the main idea" would improve reading scores more effectively than learning about ancient Egypt or the solar system or the reasons why Nevada has just as many senators as New York.

By an iron law of unintended consequences, the low scores of seventeen-year olds were probably caused by misguided, time-consuming efforts to raise scores. This chapter will predict that the high-stakes reading tests that accompany the new Common Core State Standards are going to have a similar nugatory or depressing effect on the reading competence of our seventeen-year-olds of the future—unless we take strong steps to make knowledge acquisition a chief goal of schooling starting in the earliest grades.

HOW CURRENT HIGH-STAKES READING TESTS AFFECT SCHOOLING

> *Till that which suits a part infects the whole,*
> *And now is almost grown the habit of my soul.*[2]

As I write now, the too-frequent testing of students in American schools has at last become a subject of concern and

even self-criticism by the US Secretary of Education and the President of the United States.[3] An October 2015 report from the US Department of education states: "Done poorly, in excess, or without clear purpose, [tests] take valuable time away from teaching and learning, draining creative approaches from our classrooms. In the vital effort to ensure that all students in America are achieving at high levels, it is essential to ensure that tests are fair, are of high quality, [and] take up the minimum necessary time."[4]

Secretary Duncan and President Obama had probably been prompted by an alarming study recently produced by the Council of Great City Schools reporting that each student in the large cities now takes about eight standardized tests in a year. Just sitting for the tests takes over 4 percent of school time.[5] But that's only the tip of the iceberg. The school time usurped by sitting for the tests has not been the most disabling consequence of over-testing.

In language arts, for instance, schooling has been bent out of shape and made boring and ineffective by test-prep exercises in skills like "summarizing" and "questioning the author." Yearly focusing on these exercises is supposed to improve students' comprehension of any text. And indeed students do show an *initial* positive effect from practicing finding the main idea. But their progress quickly reaches a limit and then halts. We know this from various metastudies as well as from the stagnant NAEP data.[6] Drills in formal comprehension skills have not raised mature reading scores; rather, they have taken up a lot of class time that could have been devoted to knowledge building.

Nervous superintendents and principals have insisted on

these test-prep programs, acting on the theory that strategies are the keys to reading comprehension. Teachers have been told that subject matter is secondary, that they can teach strategies just as well with *Tyler Makes Pancakes!* or *Stupendous Sports Stadiums* as with a biography of Abraham Lincoln. This emphasis on technique at the expense of building subject-matter knowledge in early grades produces students who at age seventeen lack the knowledge and vocabulary to understand the mature language of newspapers, textbooks, and political speeches. They can read *Stupendous Sports Stadiums*, but not *The Economic Consequences of the Peace*.

How can a teacher know whether too much time is being spent on practicing formal skills like summarizing? The basic principle is straightforward. We know that the skill of reading comprehension in any given case depends more on relevant knowledge than on formal strategies.[7] Once briefly learned, the strategies will take care of themselves.[8] When a school follows a coherent and specific knowledge-based curriculum, and makes that knowledge the chief object of tests, then we can make sure that students are advancing in the substantive knowledge set forth in that unit. The basic principle for schools to keep in mind is that, once decoding has been mastered, the "skill set" that most reliably determines reading comprehension is relevant knowledge. The wise teacher and school will therefore create better summarizers and main-idea finders automatically if they focus on knowledge building—a happier, more productive, and far less boring focus for schooling.

TECHNICALLY VALID, EDUCATIONALLY INVALID

READING TESTS

There is no way of predicting the topics that will appear in passages on current reading tests. That is an expected, even obvious, characteristic of an all-purpose sort of reading test, which is, of course, the only sort that could possibly be fair when the topics of the school curriculum are unknown. That structure automatically forces schools to focus on strategies and skill drills, rather than on systematic knowledge acquisition. The unpredictability of the test topics forces the schools to stress the externalities of test taking and of meaning guessing.

Despite their indifference to the school curriculum, the better reading tests are technically reliable and valid. The Gates-MacGinitie reading test, the Iowa Test of Basic Skills, and the National Assessment of Educational Progress are good measures of the average reading abilities of large groups of students. Gates and Iowa each show strong internal correlations between different forms of the same test (about .9 on a scale in which 1 is perfect). That means that the tests are reliable and that the same person will score about the same on different forms of the test. Well-established tests like Gates, Iowa, and NAEP are also technically valid. They really do roughly test a student's average reading ability. They show a fairly high correlation with other reliable tests (about .7 to .8).[9] All these well-calibrated tests are probing the average level of a person's reading fluency and vocabulary size. Such well-established tests are the means by which we confidently know about the average reading abilities of our students. That's how we know that our nine- and thirteen-year-olds have improved, and our seventeen-year-olds have not.

But these tests as currently used by the schools have hindered, not raised, mature reading skills. That is to say, they are technically valid but educationally invalid, a distinction brought into prominence by the important testing theorist Samuel Messick.[10] His phrase was "consequential validity." His key insight was that the technical validity of a test is of little value by itself. The point of educational testing is to help education. If a test actually hurts education, then it is ultimately *invalid* for schooling. Current reading tests, by giving the misleading appearance that they are testing generalized how-to skills that don't exist, cause schools to engage in self-defeating practices. They are consequentially invalid.

That defect lies less in the tests themselves than in the scientific shortcomings of the empty state standards on which they are based. The standards have misled the schools regarding the nature of reading skill. By focusing on main-idea finding, the standards promote the myth that there is a generalized main-idea-finding skill, which if practiced and developed will enhance reading ability. But if that were so for mature reading ability, the current generation of students would be performing better on the tests, for they have all been well schooled in main-idea finding.[11]

The test questions about main ideas and inference making imply the misleading message that they are probing all-purpose strategies and skills of predicting, summarizing, and "inferencing." But they are doing no such thing. The tests are probing knowledge and vocabulary. To the credit of the No Child Left Behind Act, the new focus on decoding has been highly productive for the mechanics of early reading. But

once decoding has been mastered and fluency attained, relevant knowledge becomes the chief component of reading skill. Every cognitive scientist specializing in the subject would agree with that statement.[12] No doubt unintentionally, and with inadequate knowledge of psycholinguistics, the test makers are implying a lie. By the form of their questions they suggest that they are probing formal skills. But, no matter how well trained students become in main-idea finding, the student with the smaller relevant vocabulary and knowledge is the one who will fare worse on the test.

That parents and teachers alike are demonstrating against the new tests, and opting out of them, is a pretty good indication that they have correctly concluded that something has gone wrong. The opt-out protest against excessive testing and test preparation has spread to significant numbers of districts.[13] Teachers have complained that test preparation has narrowed elementary schooling—pushing out social studies, science, and the arts. Parents have complained of the neglect of history and the arts, adding that constant testing has placed unfruitful stress upon their children. This from a recent story in the *New York Times*:

> Parents railed at a system that they said was overrun by new tests coming from all levels—district, state and federal. Some wept as they described teenagers who take Xanax to cope with test stress, children who refuse to go to school and teachers who retire rather than promote a culture that seems to value testing over learning.
>
> "My third grader loves school, but I can't get her out of the car this year," Dawn LaBorde, who has three children in Palm Beach County schools, told the gathering, through tears. Her son, a junior, is so shaken, she said, "I have had to take him to his doctor." She added: "He can't sleep, but he's tired. He can't eat, but he's hungry."

One father broke down as he said he planned to pull his second grader from school. "Teaching to a test is destroying our society," he said.[14]

Such protests need to be channeled into productive directions. They can be.

DEFECTIVE READING STANDARDS

State tests in math have been based on specific content standards, but the situation is vastly different in reading, where test makers in their own defense can rightly say: "How is it possible to create a test that encourages the imparting of concrete knowledge when the *standards* on which the tests must be based are content free and encourage the teaching and testing of general skills?"

The makers of standards have decided that while it is politically feasible to create a definite content guide in math, fierce controversy would follow if they created a definite content guide in reading. So, American makers of standards have felt themselves forced into content cop-outs in reading. Current math standards are much better guides for teachers and test makers than are current standards in reading.

I'll illustrate this with a few examples comparing the two kinds of standards. Here are some current Texas math standards concerning fractions. They are clear and content-specific, and build on one another from year to year.[15]

Grade 3: "Explain that two fractions are equivalent if and only if they are both represented by the same point on the number line or represent the same portion of a same size whole for an area model."

Grade 4: "Represent a fraction a/b as a sum of fractions 1/b, where a and b are whole numbers and b > 0, including when a > b."

Grade 5: "Represent and solve addition and subtraction of fractions with unequal denominators referring to the same whole using objects and pictorial models and properties of operations."

Here are some Texas reading standards concerning informational texts. They are neither specific nor progressive. Wary of *specifying* either topics or texts, the standards makers focus on the skill of finding the main idea.

Grade 3: "(A) identify the details or facts that support the main idea."

Grade 4: "(A) summarize the main idea and supporting details in text in ways that maintain meaning."

Grade 5: "(A) summarize the main ideas and supporting details in a text in ways that maintain meaning and logical order."

Grade 6: "(A) summarize the main ideas and supporting details in text, demonstrating an understanding that a summary does not include opinions."

Grade 7: "(A) evaluate a summary of the original text for accuracy of the main ideas, supporting details, and overall meaning."

Grade 8: "(A) summarize the main ideas, supporting details, and relationships among ideas in text succinctly in ways that maintain meaning and logical order."

My own state of Virginia is more forthright about the inherent repetitiousness and content emptiness of its reading standards:

Grade 2: "g) Identify the main idea."

Grade 3: "g) Identify the main idea."

Grade 4: "d) Identify the main idea."

Grade 5: "g) Identify the main idea."

Grade 6: "g) Identify the main idea."

Grade 7: "g) Identify the main idea."

Grade 8: "h) Identify the main idea."

Nor are the new Common Core State Standards in English language arts exempt from this same lack of real progression or content:

Grade 3: "Determine the main idea of a text; recount the key details and explain how they support the main idea."

Grade 4: "Determine the main idea of a text and explain how it is supported by key details; summarize the text."

Grade 5: "Determine two or more main ideas of a text and explain how they are supported by key details; summarize the text."

Grade 6: "Determine a central idea of a text and how it is conveyed through particular details; provide a summary of the text distinct from personal opinions or judgments."

Grade 7: "Determine two or more central ideas in a text and analyze their development over the course of the text; provide an objective summary of the text."

Grade 8: "Determine a central idea of a text and analyze its development over the course of the text, including its relationship to supporting ideas; provide an objective summary of the text."

Such standards are not just empty; they are deeply flawed. The notion that skill in finding the main idea can take the place of content is worse than empty; it's actively misleading. There is *no* reliable main-idea-finding skill. If readers understand a passage, they will reliably answer the test question about the main idea. If they don't understand the passage, they won't. Moreover, in good, complex writing there isn't usually a single main idea. What's the main idea of the Pledge of Allegiance? Aren't there at least three?

These empty standards were created out of political expediency. The makers of standards and tests have built up

an artificial construct, politically painless for the makers of standards and of tests, but based on a faulty and unproductive picture of reading comprehension.

EMPTY SKILL STANDARDS CAUSE TESTS TO PRETEND TO PROBE EMPTY SKILLS

Test makers have dutifully followed the standards makers, presenting reading comprehension as an all-purpose skill like decoding. Since it isn't, the current standards and tests have created a fictional alternative universe in our classrooms. In the real world, an ability to comprehend a piece of writing depends on one's having the knowledge and vocabulary relevant to that passage. If the school does not teach students the knowledge and vocabulary they need to understand the passages on the test, then the test is unfair as a measure of what the school has successfully taught. (Yes, that means that the current tests *are* unfair, a subject I take up in more detail in the next chapter.) The test simply reflects the knowledge and vocabulary that students have picked up from *all* sources. A school test that does not accurately measure the matter that a student has been taught in school is an unfair test of schooling. Such tests cannot measure whether students have mastered the knowledge and vocabulary that the school *has* taught. Test makers cannot know what knowledge and vocabulary schools have taught. The standards do not state them. Nor do the schools know what knowledge they are supposed to teach. The language arts standards do not specify content.

Under these circumstances, a method has had to be devised that *seems* to make these inherently unfair tests fair. That

method has been to define reading ability as a set of strategies, and then to create test items that appear to probe those strategies. The external forms of the test questions are constructed to give the impression that they are testing the various skills that were being practiced so endlessly in test-prep classes. Here are some "stems" taken from released items on the Texas tests. Their form indicates misleadingly that strategy expertise rather than specific knowledge is being probed.

The main purpose of paragraphs 7 and 8 is to [main idea]
The author wrote this selection most likely to tell the reader that [main idea]
Which sentence expresses the main idea of paragraph 2?
The reader can infer that Chu is concerned about Dusty's habit because she ["inferencing skills"]—
Which of these best summarizes the selection? [main idea]
The reader can infer that the author's attitude toward Dusty is one of [inferencing skills]
The reader can infer that the long life span of bristlecone pines is mainly a result of
What is the main idea about bristlecone pine trees presented in the selection?
The organization of paragraphs 2 through 4 contributes to the author's main idea by[16]

Answering such items is easy for students who understand the passages, but not for those who don't—no matter how many drills in comprehension strategies the students had before the test. A child who has the relevant domain-specific background knowledge will understand the passage and get the right answer fast, without conscious strategizing. A child who does not have enough relevant knowledge will have to use special glosses in the test and consciously apply

strategies; that child won't finish the test and will get many answers wrong.

In sum, to mask the inherent unfairness of these tests, a fictitious alternative world has been devised in which metaskills look as important as knowledge and vocabulary. The substance of the reading curriculum had been described in the content-evading state standards as consisting of all-purpose techniques like inference making, predicting, and main-idea finding.[17] With such standards, all-purpose techniques *became* the curriculum, and reading tests that asked main-idea questions seemed to be "standards-based" and "curriculum-based." Despite their unfairness (especially to disadvantaged students), they have thus been made to appear to be fair.

But aren't there in fact such all-purpose strategies? And won't learning them improve reading? The evidence is summarized in a recent review article by Willingham and Lovette, "Can Reading Comprehension Be Taught?"[18] They answer: "Not really." Lessons in reading strategies offer an initial score boost for test taking, but are quickly learned; they plateau fast, and they don't have to be practiced. Their utility ceases after about ten lessons. Two weeks on comprehension strategies is optimal. There is no practical utility after that. Huge amounts of time are being wasted. Worse, making young students become highly self-conscious about applying strategies distracts their attention and degrades their performance.[19]

The "accountability" principles based on this misconceived scheme have not induced real progress in higher-level reading competence. If progress is to be made, an alternative

structure of teaching and testing reading will have to be instituted. The structure will need to become more like that of math, with specific content in the standards and the curriculum, and with tests based on that content. Perhaps the Common Core State Standards will ultimately move in that direction.

THE NEW COMMON CORE READING TESTS

I will devote a later chapter to the Common Core initiative. The responses of many schools to the new Common Core tests can already be seen. Many school administrators are responding to them as unproductively as they did to prior high-stakes tests under No Child Left Behind. Schools are intensively practicing techniques like making inferences and finding the main idea, and now they are also practicing close reading and complexity managing in preparation for the new versions of high-stakes reading tests that, as before, pretend to test the general skills of close reading, complexity managing, and main-idea finding—general skills that do not exist.[20]

It's with some reluctance that I end this chapter with a criticism of the reading tests being designed by the test consortia for the Common Core State Standards. The Common Core Standards themselves are in some respects superior to most standards for individual states. They contain the welcome admonition that the standards can only be properly implemented through a "curriculum intentionally and coherently structured to develop rich content knowledge within and across grades."

Those precious words are to my mind the most important

ones in the new standards. But there is little sign that districts or states are paying much attention to them, for that general admonition is not a standard that can readily be coordinated with the items on a reading test. In the absence of specific grade-by-grade content guidance, the makers of the Common Core tests are placed in the same position as the makers of the previous tests that were based on state standards. The new test items will need to probe main-idea-finding skills and inference-making skills, as before, and now they will probe close-reading skills as well.

So, despite the expense and computerized novelty of the new reading tests, it will be hard for them to be any more educationally productive than the tests they are displacing. The new reading tests, like the old ones, will need to be based upon main ideas and inference making. The new kind of inference is to be connected with an additional inference strategy called "close reading," according to the standard: "Read closely to determine what the text says explicitly and to make logical inferences from it."[21]

The sample items made public by the new test consortia for the Common Core amply fulfill my prediction that the tests will stress both main-idea finding and close reading. Those are the Common Core standards that are the most content-free, strategy-like elements. Thus PARCC, one of the two test consortia, has developed two-part questions, the first part gauging comprehension of a passage, the second part demanding a close reading and "logical inference" to justify the student's answer. Here's a multiple-choice example from PARCC for grade 3:

Part A: What is the meaning of the word "avenge" as it is used in

the story?

Part B: Which detail from the story best supports the answer to Part A?

This structure is perfectly reasonable, so long as schools can be brought to realize that the secret to answering such questions will not be hours of practice of "inferencing skills" and "close-reading skills," but can only be answered through the student's prior relevant knowledge of the words and the topics.

The "Smarter Balanced" consortium puts the two-part structure into a single question, such as this one for grade 4:

Read the sentence from the text. Then answer the question.

"Nanodiamonds are stardust, created when ancient stars exploded long ago, disgorging their remaining elements into space."

Based on the context of the sentence, what is the most precise meaning of disgorging?

scattering randomly
throwing out quickly
spreading out widely
casting forth violently

No doubt, the student is meant to answer the question in the following way after close reading: "Let's see. The stars exploded—a violent event. *Disgorge* must mean that the bits were cast forth violently. So 'casting forth violently' must be right."

But this very example illustrates the inadequacy of the suggestion that close reading consists in "making logical inferences." Readers who already know the word *disgorge* know that the sense of the word can make *all* the proffered meanings correct—that the nanodiamonds were "thrown out

quickly" or "spreading out widely" or "scattering randomly," in addition to the supposedly correct answer. The right answer depends on what we decide the passage is meant to emphasize: whether the result of the action or the action itself. I'd guess that the author meant to imply something not all that violent, since *disgorge* is usually not used to describe violent, explosive action. Anyway, logical acumen will not decide that particular issue. Either full credit should be given to *all* the nanodiamond answers, or the question should simply be nullified. It's based on a wrong theory of inference and reading comprehension.[22]

Much more interesting is the phrase "remaining elements" in the passage. The phrase is ignored in the question, and is far less susceptible to close reading and logical analysis. One could close-read "remaining elements" all day long and come up empty or wrong. Its meaning requires specific background knowledge of a highly sophisticated kind (probably already provided earlier in the source), which most fifth graders will normally lack. It's conveniently ignored in the test item, but well illustrates how knowledge trumps "skill" every time. Knowledge is by far the most promising avenue to carry us out of the reading slump we are in. It is by far the most promising way to advance reading skill for all, and narrow the reading gap between demographic groups.

I recently did a *Huffington Post* piece with the following subtitle: "The Common Core Tests in Language Arts Will Soon Be Coming to Your Child's School. Tell Your Local Superintendent: 'Don't Worry. Students Will Ace Those Tests If They Learn History, Civics, Literature, Science, and the Fine Arts.'"[23] The comments by teachers on my piece were

sobering. They carried this message: "Well, Mr. Hirsch, that's all very well, but if you were in my shoes you'd realize that your job would be at stake if you did not do test preparation as instructed."

Quite right.

Parents and teachers need to get vigorously involved in the testing issue, not just to complain about stress, but to change the character of reading tests and reading instruction. Those wasted hours ought to be spent on far more interesting and rewarding subject matters that will build up knowledge and vocabulary, and therefore induce greater reading competence. A positive aim of parents should be to demand knowledge-building substance in their child's language arts classroom, to replace exercises that are knowledge-displacing, soul-deadening, and essentially useless after ten lessons. When a student gains a real advance in substantive knowledge, it's the best long-range comprehension strategy.

Yes, we do need tests. Yes, students, teachers, and schools should be accountable. But accountable for what? The standards have not told them with adequate specificity or adequate insight into the actual nature of reading. Only standards that are guides to curricular content can be foundations for fair and productive tests. Policy makers who stress accountability are right to do so. But they must come to understand that without definite content standards, there can be no fair and productive accountability in reading. The authorities can't just wave their hands and imperiously demand better reading without stating what children need to learn and schools need to teach in a given grade in order to gradually reach that long-range goal.

Accountability hawks need to show more grit in creating state and local curriculum standards. They should join with parents and teachers to help define the grade-by-grade knowledge that all the children in their local purview should gain. A more responsible view of accountability would recognize that providing definite content standards is the only policy that can possibly lead to productive and fair reading tests.

Meanwhile, a positive intermediate step would be to remove the punitive threats to teachers attached to our educationally invalid reading tests—the subject of the next chapter. If the high stakes were removed, sensible teachers and principals would be willing to pay more attention to the long arc of knowledge acquisition, which is the route to producing good readers and competent high school graduates. To sum up: This chapter has shown that recent, technically valid reading tests make fraudulent claims if they are used to gauge what the schools have taught. And they have had a disastrously narrowing effect on schooling. They are "consequentially invalid." They do more harm than good. Their defects could be repaired if the tests were to be based on good, knowledge-based standards. Such reconstituted standards and tests would do far more good than harm—an outcome that will require greater courage and scientific insight than has been shown in the recent past.

CHAPTER TWO

The Scapegoating of Teachers

TEACHER QUALITY OR THEORY QUALITY?

People who emphasize teaching quality and the central importance of teachers are right to do so. Where some go wrong is in thinking that teacher quality is an innate characteristic. The effectiveness of a teacher is not some inherent competence, as the phrase *teacher quality* suggests. Teacher effectiveness is contextual. I have witnessed over and over that in a coherent school most teachers can become highly effective. My defense of teachers in this short chapter is not a defense of irresponsible, lazy, or nonperforming teachers. Like most people I am opposed to any policy that would impede the dismissal of demonstrably nonperforming teachers. Children and the community come first. Most teachers agree.[1]

Why has the topic of teacher quality suddenly reached such a crescendo? Education reform has been on the national agenda since 1983, the year of *A Nation at Risk*. Only in the last few years has the teacher quality issue risen to the top. I think it may be reform fatigue, possibly desperation. We are blaming teachers because of our disappointments with the results of our reforms.

The "back-to-basics" and "whole-school reform" strategies disappointed. The state standards movement and the No

Child Left Behind law have left high school students just about as far behind as they were before the law was instituted. Charter schools, despite their laudable triumphs, are highly uneven in quality.[2] Their overall results are not much better than those of regular schools.[3] When favored educational ideas do not pan out as hoped, reformers understandably think: "The flaw is not in my theory; it must lie in poor implementation (i.e., it must be the fault of the teachers)."

But the most likely cause of disappointing results from the various reforms is that they have been primarily structural in character. They have not systematically grappled with the grade-by-grade specifics and coherence of the elementary school curriculum. Educational success is defined by what students learn—the received curriculum. Not to focus on the particulars of the very thing itself has been an evasion that is not of the teachers' doing. The underlying theory of the reforms (reflected in state reading standards) has been that schools are teaching skills that can be developed by any suitable content. That mistaken theory has allowed the problem of grade-by-grade content to be evaded. It was that fundamental mistake about skills that has allowed teachers to be blamed for fundamental failures—the failures of guiding ideas, not of teachers.

Elementary school teachers are people who for the most part love children, who want to devote their lives to children's education, but find themselves stymied and frustrated in the classroom. They apply the notions received in their training, and do what they are told to do by their administrators, under the ever-present threat of reading tests

that do not actually test the content that is being taught. Under these extremely unfavorable conditions of work, it's no wonder that teacher unions have focused on bread-and-butter concessions—and have pushed back against punitive but unproductive reforms. When the classroom, which should be a daily reward, becomes a purgatory, one turns to contract stipulations.

It's true that in the United States, there has been a deep problem with teacher preparation for more than half a century. We have a system that, according to teachers themselves, does not prepare them adequately for classroom management or the substance of what they must teach.[4] Yet, as I will illustrate with the example of France, even with a staff of well-trained, highly qualified, and well-educated teachers, the schools can suddenly decline when a substantive curriculum is abandoned, and fallacious ideas about skills begin to dominate. Therefore my counterthesis to the blame-the-teachers theme is blame the ideas—and improve them.

When an early Core Knowledge school was started in Fort Myers, Florida, in 1990—the Three Oaks School—I visited the school and its upbeat, excited teachers. They were intimidated that first year by having to teach things they did not know themselves. The next year when I visited, they were enthusiastically explaining that they were learning these things along with the children. They started having a signal success in improved morale of students and teachers, and improved test outcomes. If "teacher quality" is to be judged by outcomes, their quality had suddenly risen.

The "quality" of a teacher is not a permanent given. Within

the content-incoherent American primary school, it is impossible for a superb teacher to be as effective as a merely average teacher is in the content-cumulative Japanese elementary school. For one thing, the American teacher has to deal with big discrepancies in student academic preparation, while the Japanese teacher does not. In a system with a specific and coherent curriculum, the work of each teacher builds on the work of teachers who came before. The three Cs—cooperation, coherence, and cumulativeness—yield a bigger boost than the most brilliant efforts of teachers working individually against the odds within a topic-incoherent system. A more coherent system makes teachers better individually and hugely better collectively.

American teachers (along with their students) are, in short, the tragic victims of inadequate theories. They are being blamed for intellectual failings that permeate the system within which they must work. The real problem is idea quality, not teacher quality. The difficulty lies not with the inherent abilities of teachers but with the theories that have watered down their training, and created an intellectually chaotic school environment based on developmentalism, individualism, and the skills delusion. The complaint that teachers do not know their subject matter would change almost overnight with a more specific curriculum and with less evasion about what the subject matter of the curriculum ought to be. Then teachers could prepare themselves more effectively, and teacher training could ensure that teacher candidates have mastered the content they will be responsible for teaching.

Those who hope to find amelioration of the "teacher

quality problem" through the use of computers and "blended learning" may be fostering yet another skills delusion. Technological fixes haven't worked in the past. Computers seem to work best in helping older students learn specific routines. No doubt well-thought-out computer programs can help teachers do their work, especially for teachers in their first years. But there are *inherent* limitations. For example, after decades of work and billions spent, computers cannot accurately translate from one language to another. Probably they can't even in theory.[5]

Such current limitations do not lend confidence that they can transform primary education. Young students rely on an empathetic personal connection that not even our most advanced computer-adaptive programs can deliver. This is not to say that computers have no important place; it is to say that their place is supplemental, not transformative. They need to be used in support of teachers under a coherent cumulative curriculum. Computers cannot magically replace the hard thinking and political courage needed to create one.

A FATAL FLAW IN VALUE-ADDED TEACHER EVALUATION

In the face of unfair scapegoating, teachers have understandably become demoralized by being constantly blamed for failures not their own. Here is the new conventional wisdom about teachers taken from the nonpartisan policy magazine *Governing* of June 13, 2013:

> The research is clear: Teacher quality affects student learning more than any other school-based variable (issues such as income and parental education levels are external). And the impact of student

> achievement on economic competitiveness is equally clear. That's why it's so disturbing that in 2010, the SAT scores of students intending to pursue undergraduate education degrees ranked 25th out of 29 majors generally associated with four-year degree programs. The test scores of students seeking to enter graduate education programs are similarly low and, on average, undergraduate education majors score even lower than the graduate education applicant pool as a whole. Education schools long have accepted under-qualified students, then offered them programs heavy on pedagogy and child development and light on subject-matter content.

This scientific-sounding comment is incorrect from the start. The assertion that "Teacher quality affects student learning more than any other school-based variable" is not footnoted. According to two summaries of research by Dr. Russ Whitehurst, a better curriculum can range from being slightly to dramatically more effective than a better teacher.[6] That's not surprising when you consider that the curriculum is what teachers teach and what students are supposed to learn. Teachers are not to blame for ideas and curricula that are inherently inadequate.

Some policy makers have recently decided that the way to improve teacher effectiveness is to institute value-added teacher evaluations as part of a system of incentives, rewards, and sanctions, potentially including dismissal. The theory is that such a system will energize teachers, boost their performance, and bring highly qualified people into the profession. Some jurisdictions, including Chicago, Washington, DC, and New York City, have instituted value-added measures (VAMs) of teacher effectiveness, based on formulas like:

$$Ag = \theta\, Ag{-}1 + \tau j + S\phi + X\gamma + \varepsilon$$

where Ag is the achievement of student i in grade g (the subscript i is suppressed throughout); Ag–1 is the prior year student achievement in grade g – 1; S is a vector of school and peer factors; X is a vector of family and neighborhood inputs; θ, ϕ, and γ are unknown parameters; ε is a stochastic term representing unmeasured influences; and τj is a teacher fixed effect that provides a measure of teacher value added for teacher j.[7]

Statistical analysis is indispensable, but can be very misleading unless supported by a valid theory of the underlying causes of the results. But, in fact, the results themselves cry out that something is amiss, since the value-added principle has exhibited far more uncertainty and variability for language arts than for math. That's not surprising. In math, as I showed in chapter 1, there is a high correlation between what is supposed to be taught and what is actually tested, whereas that's not true for the language arts curriculum and current reading tests.

Two false assumptions underlie applying VAMs to reading tests. The first mistake is the assumption that reading comprehension is a general skill. The second is the assumption that existing reading tests can accurately gauge the value that has been added by the teacher to reading comprehension from one year to the next. Our current reading tests cannot in fact reliably and validly gauge the value the teacher has added.

Here's why. Scores on reading tests reflect knowledge and vocabulary gained from all sources. Advantaged students are

constantly building up academic knowledge from both inside and outside the school. Disadvantaged students gain their academic knowledge mainly inside school, so they are gaining less academic knowledge overall during the year, even when the teacher is conveying the curriculum effectively. This lack of gain outside the school reduces the chance of low-socioeconomic-status (SES) students showing a match between the knowledge they gained in school during the year and the knowledge required to understand the individual test passages.[8] The tests are fairly accurate means of gauging a student's general knowledge, but they have no way of indicating the sources of students' general knowledge. Not being curriculum based, they cannot be an accurate means of testing how well the particular knowledge in the school curriculum has been imparted. The implicit assumption that "general reading skill" is itself the content of the curriculum is a technical mistake and an incorrect assumption. Once that mistake has been exposed, the validity of the VAM projects in language arts collapses. Any judge in a lawsuit, properly alerted to the falsity of their assumptions, should rule against the fairness of value-added measures for rating language arts teachers. These reading tests may be roughly accurate measures of a student's average reading abilities, but, not being curriculum based, they *cannot* be accurate measures of school-driven gains in a given year.

In short, there's no valid or reliable way of determining what test-relevant verbal knowledge is school based and what is not. How could it be determined? *Tests that are curriculum-blind cannot gauge how well a curriculum has been imparted.* VAMs in reading are thus inherently unfair both to low-SES

students and to their teachers. Reading tests at best are 70 percent accurate at the individual level.[9] The inherent uncertainty of the school-based contribution to a student's reading scores between one year and another must reduce the validity of test inferences even more. Statistical manipulations cannot make a test reveal what it cannot reveal in principle. The whole VAM effort in reading will need to meet this objection head-on in order to establish the effort's validity. It's hard to see how it could do so. It has not done so thus far.

Another evil consequence of test-based evaluations of language arts teachers has been the demoralization of millions of public school teachers.[10] But merely avoiding unfairness to teachers does not solve the underlying problem, which is rooted in incorrect ideas that teachers themselves have often not cast aside. Like their administrators, they have been indoctrinated in individualistic, child-centered education. My plea to teachers—for the sake of their students, and themselves—is to rebel against the skills delusion; to insist on coherent and cumulative multiyear content; then cooperate and consult.

That teachers cannot be replaced by computers doesn't mean that individual teachers should not be replaced. The problems with teacher evaluations that I have discussed in this chapter concern the unreliability of the value-added measures of teacher performance in language arts, but do not apply to estimates of poor teaching based on clear evidence. There is no reason that teachers should enjoy special job protections at the expense of children. Tenure protections at universities were instituted to avoid censorship of opinion.

But even in universities there is no tenure protection for "failure to meet a specified norm of performance or productivity"; nor should there be for schoolteachers.[11]

If I were a principal in a primary school I'd spend my money on teachers, on their ongoing development, and on creating conditions in which the work of teachers in one grade supports the work of teachers in the next, and in which teachers would have time to consult and collaboratively plan. One especially vivid story about collaboration in the Japanese elementary school was told to me directly by the late Professor Harold Stevenson, who studied Asian schools. He had observed the event in a fourth-grade math class. A student was having grave difficulty with a math problem and its concepts. After allowing the student to work on it for a short time, the teacher quietly made a surprising analogy with the student's daily experience as a way of dealing with the problem. The student's face brightened, and he instantly began to solve the problem.

After the class, Stevenson went to the teacher to congratulate her (in perfect Japanese) on the most remarkable bit of teaching he'd ever witnessed. The teacher shook her head: no, it wasn't her brilliance that produced the result, and from her desk drawer she took out a handbook that teachers had cooperatively compiled. "Here it is," she said. "It's suggested as a good tack to try when you run into that situation." The incident illustrated how good teaching can often depend more reliably on the coherence of the wider system, and the cooperation it brings, than on virtuoso performances. Schooling takes twelve years. Its success depends on slow but sure progress, not bursts of brilliance—


image
not
available

vocabulary size, or academic achievement between disadvantaged children who attended Head Start and disadvantaged children who did not.[2] Similar null results were recently reported in Georgia and Oklahoma.[3] Fadeout is usually explained by the observation that out-of-school life experiences of disadvantaged children have drowned out their in-school experiences—a self-evident truth.

But that explanation does not tell us why some American elementary schools and some whole nations have been able to overcome preschool fadeout. In them, Amana keeps her initial head start and extends it. One finds in the United States happy exceptions to fadeout.[4] Add to them the vast natural experiment recorded in detail in France between 1987 and 2012. In the 1980s, research proved that French preschools had not only avoided fadeout but by age ten had greatly narrowed the achievement gap.[5] But after the radical *loi Jospin* ("Jospin law," described in chapter 7) went into effect in 1990, the excellent French preschools suddenly ceased narrowing the achievement gap. Preschool fadeout had come to France.[6] The excellent preschools had not changed; only the elementary schools had. That is a crucial piece of evidence in the preschool fadeout puzzle.

Another piece of evidence about the critical importance of the primary grades for disadvantaged students is a second fadeout phenomenon that occurs in later grades as students approach high school. Chapter 1 showed that in recent years, the whole cohort of students who have improved their verbal scores at age thirteen show no advance over prior years by the time they reach age seventeen. The progress they made as thirteen-year-olds has disappeared. So there is fadeout after

INDEX