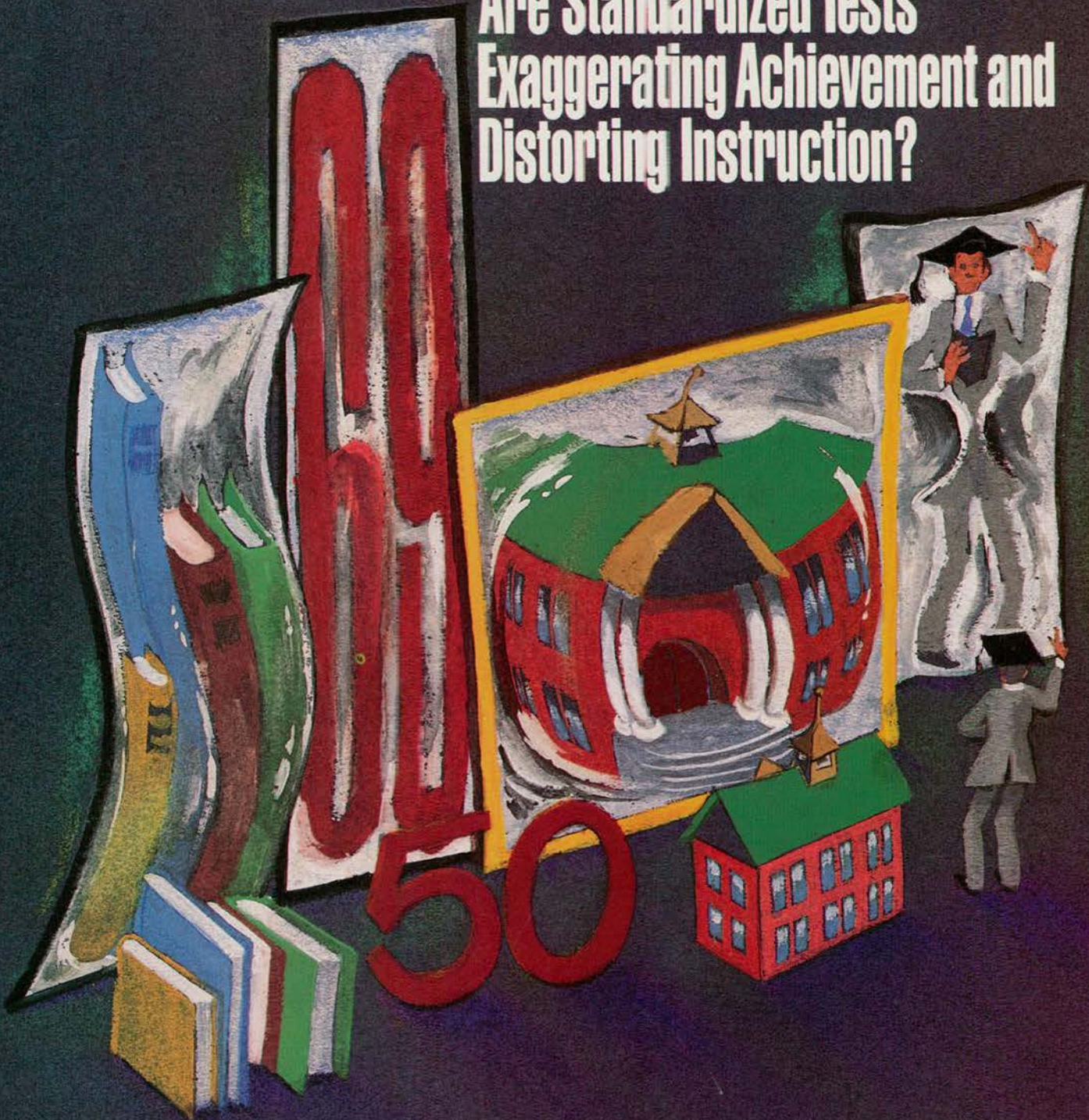


AMERICAN

AMERICAN FEDERATION OF TEACHERS
SUMMER 1988

Educator

Are Standardized Tests
Exaggerating Achievement and
Distorting Instruction?



R. Barkin



HAS TAX REFORM PLACED YOUR TAX DEDUCTIONS IN CHECK?

VALIC Tax-Deferred Annuities give educators the opportunity to offset the many tax deductions no longer available as a result of the Tax Reform Act of 1986. Deductions that have been eliminated or severely restricted include:

Sales tax deductions • Two-earner deduction • Non-reimbursed business deductions • Consumer loan interest • Capital gains exclusion allowance • Investment tax credits • Political contribution tax credits • IRA deductions subject to certain income level restrictions • And all but the most serious medical expenses.

With a VALIC TDA, you may be able to reduce your taxable income by up to 20% of your salary, which with elective contributions can total up to \$9,500. You can select from variable options with the earning power of mutual funds or from fixed options with attractive, fixed interest rates. VALIC's tax-free loan provision allows you access to your TDA assets without the federal tax penalties normally associated with early withdrawal.

It's your move. Return the coupon below and discover how VALIC can help you plan for retirement.

Please have a VALIC Representative contact me:

Name _____

Address _____

City _____

State _____

Zip _____

Home Phone _____

Work Phone _____

Employer _____

AE

Send to: VALIC Group
Marketing, 2929 Allen
Parkway, 7th Floor A7-25,
Houston, Texas 77019

 **VALIC**

★ An American General Company

We'll help you shoot down high textbook costs...



Purchasing Used Books
SAVES $\frac{1}{3}$ TO $\frac{1}{2}$

ELEMENTARY AND HIGH SCHOOL TEXTBOOKS AND WORKBOOKS
ALL PUBLISHERS • USED, REBOUND, AND NEW • CURRENT AND OLDER EDITIONS



WILCOX & FOLLETT BOOK COMPANY
1000 West Washington Blvd.
Chicago, Illinois 60607
1-800-621-4272

(Circle below, call, or write for our FREE catalog)

For your next IBM presents its

The newest member of the IBM Personal System/2 family was made for a very special desktop: the one in the classroom.

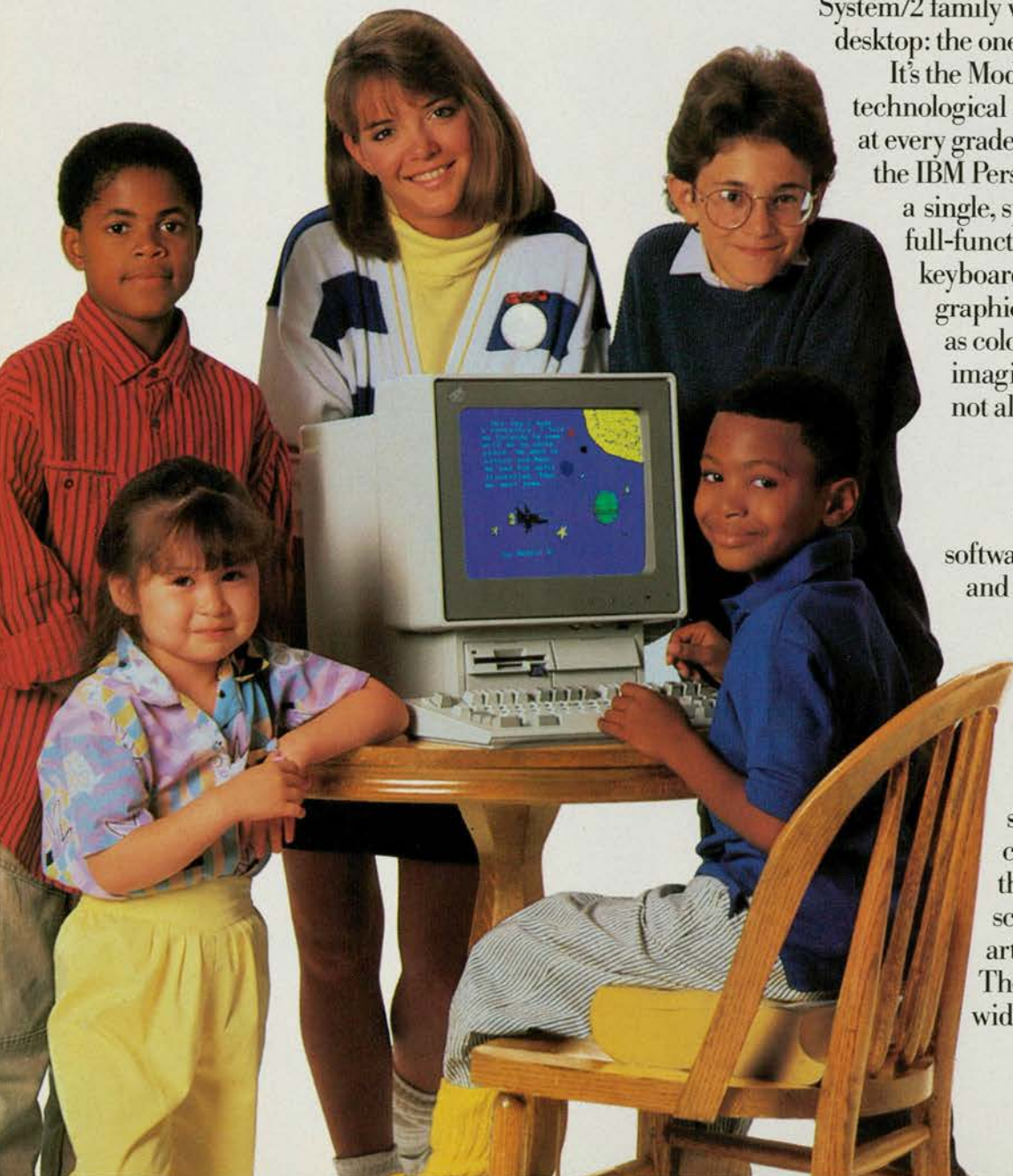
It's the Model 25. It brings our newest technological advancements to students at every grade level. It packs the power of the IBM Personal System/2 family into a single, space-saving unit. It has a full-function, easy-to-understand keyboard, as well as advanced graphics and animation that are as colorful as a student's imagination. Of course, that's not all it has.

IBM courseware.

The IBM educational software library was designed and developed by educators.

Furthermore, it was tested by a demanding group of critics: teachers and students, who gave it high marks.

We offer an extensive selection of high quality courseware that covers all the basic skills: math, science, reading, language arts and business education. The Model 25 also runs a wide selection of courseware



generation, next generation.

developed by independent software authors.

IBM networking.

Through the IBM classroom network, the teacher's computer can be linked to each



Typing for Beginners



Biology Series



Physics Series

fits nicely on a student's desk. So set-up is really a one-step process, you just plug it in and go.

We also make it easy to buy. We've reduced the price of the Model 25 for faculty and staff, plus we offer schools special prices on the IBM Personal System/2 family.

Of course, you get more than a box with IBM; you also get the support and service you'd expect from IBM. We even have a team of Education Instruction Specialists (teachers working with IBM), who offer training

and support to help you get off to a good start.

With the IBM Personal System/2 family, you have all you need to run a model school system.

If you'd like to see how easy it is to bring your next generation and our next generation together, call 1-800-IBM-2468, ext. 25. Or write to us at Dept. 25, 101 Paragon Drive, Montvale, NJ 07645, for more information.

IBM The Bigger Picture

student's. This, in turn, allows the teacher to give each student specific instructions, as well as distribute the day's work electronically. Also, the teacher knows which program a student is using and how fast they're working.

The IBM network helps students get right to work on the right subjects, and advance at their own pace.

Our network even has benefits that go beyond the classroom, all the way to the administration office. The teacher can have a direct link to the office which can make daily chores, like taking attendance, effortless.

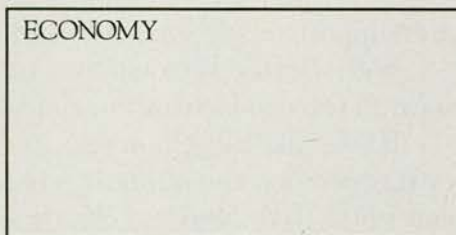
The most important result is that teachers will have more time to teach and students will have more time to learn.

IBM makes it easy.

The Model 25's disk drive, processor and monitor are built into a single compact unit that



THERE GOES THE ECONOMY.



Rent from National and economy cars aren't in the picture. Because we give a free upgrade to all AFT travellers for simply clipping this coupon. Of course, you still get your AFT discount. Because, after all, you never know what shape the economy will be in.

FREE UPGRADE

Name _____ Times in the last year you have rented a car? 1-6 ___ 7 or more ___

Address _____ City _____ State _____ Zip _____

Bearer of this certificate is entitled to a One Car Class Upgrade of an E-Economy or M-Mid-Size car at participating National locations in the U.S. Simply present this certificate at the time of rental. Certificate valid through 7/31/88.

Terms and Conditions

1. Driver must meet standard age, driver's license, and credit qualifications at time and place of rental. 2. Customer must place a reservation for the upgraded car class. 3. Certificate cannot be used in conjunction with any other certificate or special promotion. 4. Specific cars are subject to availability. 5. Offer void where prohibited by law, taxed or otherwise restricted. 6. Limit one certificate per rental. 7. Maximum rental two weeks.

Rental Agent Instructions

1. Qualify the customer for car rental in the normal fashion. 2. Verify that the car class reserved is the car class the customer actually wants and charge the rate for one car class lower. 3. Record Recap No. 1108844 and "AFT UPGRADE" on the R.A. 4. At the time of return, retain the certificate and mail to Headquarters, Attention Program Development.

 National Car Rental.

ALBERT SHANKER
President
American Federation of Teachers
Elizabeth McPike
editor

Ruth Wattenberg
associate editor
Mary Power Boyd
associate editor

Diane Aiken
copy editor

Wanda Bailey
secretary

Andrew Bornstein
design consultant

Cover illustrated by Robert Barkin

Subscriptions: AFT members receive *American Educator* as part of their AFT membership services. Nonmembers may subscribe for \$8 per year.

Signed articles and advertisements do not necessarily represent the viewpoints or policies of the American Federation of Teachers, AFL-CIO.

American Educator cannot assume responsibility for unsolicited manuscripts.

American Educator is published quarterly by the American Federation of Teachers, AFL-CIO, 555 New Jersey Ave., NW, Washington, DC 20001. Telephone: 202-879-4420.

General advertising office
555 New Jersey Ave., NW
Washington, DC 20001
Telephone: 202-879-4420

Advertising Sales Representative
Peter Li, Inc.
2451 E. River Rd.
Dayton, OH 45439
513-294-5785

William M. Doran, Michael O'Callaghan,
Frank Guttman
Peter Li, Inc.
1220 Broadway
New York, NY 10001
212-947-2300

West
Terry Casaus
The R. W. Walker Co.
2716 Ocean Park Blvd., Suite 1010
Santa Monica, CA 90405
213-450-9001

Southwest and Midwest
Thomas H. Bachmann, Patricia O'Connor,
Geoff Welch
Peter Li, Inc.
913 West Van Buren, Suite North 7B
Chicago, IL 60607
312-942-0063

American Educator is produced with the assistance of members of Local 2, Office and Professional Employees International Union, AFL-CIO, and members of AFT Staff Union. Composition and printing are done in 100 percent union shops.

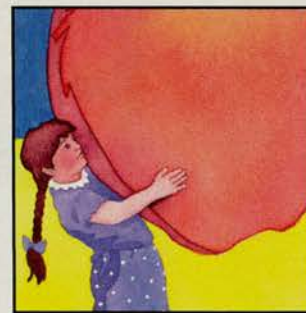
© American Federation of Teachers, 1988.



Page 16



Page 20



Page 34

AMERICAN Educator

ARRIVING IN LAKE WOBEGON: ARE STANDARDIZED TESTS EXAGGERATING ACHIEVEMENT AND DISTORTING INSTRUCTION?

8

By Daniel Koretz

Can more than half of our students be above average? A recent report revealing that standardized test scores are overstating achievement levels, often by a large margin, has led to a new, critical look at test-based accountability. In a far-reaching analysis, the author shows that a lot more is amiss than inflated scores. There are disturbing signs that the pressure to demonstrate results on standardized tests is distorting instruction and impeding learning.

THE LOSS OF JOBS AND THE RISE OF THE UNDERCLASS

16

A Review by Norman Hill

Despite gains made in civil rights, urban ghetto poverty has become more profound. The overwhelming reason, says William Julius Wilson in a new book, is that traditional working-class jobs are disappearing, leaving massive unemployment in the wake.

A CONSPIRACY OF GOOD INTENTIONS: THE TEXTBOOK FIASCO

20

By Harriet Tyson-Bernstein

Our textbooks are big, beautiful . . . and, oh, so bad. They're boring, often incoherent, and overstuffed with incomprehensible detail. How do they get this way, especially when everyone involved is trying to do his best?

MUDDLE BY MANDATE: A FICTIONALIZED ACCOUNT

21

By Harriet Tyson-Bernstein

Fourteen months from now, the State of Nirvana will adopt new textbooks. From start to finish, here's how the process goes wrong.

WHAT SHOULD YOUNG CHILDREN BE DOING?

28

By Lilian G. Katz

With the push for stepped-up academics coming from many directions, the debate over the role and functions of kindergarten is in full swing. One of the nation's foremost experts on early childhood education takes us back to the basic question that should inform all policy.

FLUNKING KINDERGARTEN: ESCALATING CURRICULUM LEAVES MANY BEHIND

34

By Lorrie A. Shepard and Mary Lee Smith

Kindergarten retention, say the authors, does nothing to boost subsequent achievement. What it does is mark with failure a child's first encounter with school and feed the pressures for an ever-more-demanding curriculum.

Face to Face...



with the Future.

The right computer lets your students see a new world of career opportunities.

Move up to the MS-DOS® standard

Your students need proper training for their future. Our line of MS-DOS computers provide that training. MS-DOS gives your students a tool for their future—a tool they'll be able to use in school, at work . . . and in life.

MS-DOS is the most popular operating system in the world today. With other operating systems, you risk incompatibility with the real world.

The Intel® 80286-based Tandy 1000 TX features one 3½" disk drive, 640K RAM, a headphone jack with volume control and Personal DeskMate™ 2 software. Adapters for a monitor, printer, joysticks and mouse are all included.



The Tandy 1000 HX features a 3½" disk drive, headphone jack with volume control and our Personal DeskMate 2 software. It also includes MS-DOS built-in—just turn it on and go!

Tandy® has the right tools

Our line of MS-DOS computers delivers the technology that today's world demands—and delivers it more affordably than IBM®. From stand-alone desktop computers to networked labs, Tandy offers the solution to your school's computer needs. And Tandy computers are backed with service and support that are unequalled in the industry.

Educational software for varied curricular needs

From kindergarten to college, quality MS-DOS educational software programs for almost any curriculum are available for Tandy computers. We offer a wide selection of educational software right off the shelf in our Radio Shack Computer Centers.

There are fun learning programs for kindergarten and elementary students, including software like Math Blaster, Where in the World is Carmen Sandiego?, Typing Tutor IV, and the award-winning Rocky's Boots, which helps develop skills in problem solving, abstract reasoning and creative thinking.

For secondary students, there's no better preparation for business than professional-level applications. Students can get an inside track on their career goals with software like Lotus 1-2-3, dBASE III Plus and WordPerfect.

Express Order Software gives you even more selection

Literally hundreds of titles are available through our exclusive Express Order Software system. These packages are as close as your nearby Radio Shack store, available in just days after your order. You'll find software from such well-known publishers as Broderbund Software, Davidson, DesignWare, The Learning Company, Scholastic, Springboard and Weekly Reader.

Total commitment to education

The Radio Shack Education Division devotes its full time to supporting Tandy computers and the teachers who use them. We never stop working to provide schools with the right tools at the right prices. Send in the coupon today for more information.

Tandy Computers:
Because there is no better value.™

Radio Shack®
The Technology Store™

A DIVISION OF TANDY CORPORATION

Send me a 1988 computer catalog.

Mail to: Radio Shack, Dept. 88-A-169
300 One Tandy Center, Fort Worth, TX 76102

Name _____
Institution _____
Address _____
City _____ State _____
Zip _____ Phone _____

Intel/Reg. TM Intel Corp. Math Blaster/TM Davidson Associates. Where in the World Carmen Sandiego?/TM Broderbund Software. Typing Tutor IV/TM Simon & Schuster. Rocky's Boots/TM The Learning Co. 1-2-3/TM Lotus Development Corp. dBASE III Plus/TM Ashton-Tate. WordPerfect/TM WordPerfect Corp. MS-DOS/Reg. TM Microsoft Corp. IBM/Reg. TM IBM Corp.

ARRIVING IN LAKE WOBEGON

Are Standardized Tests Exaggerating Achievement and Distorting Instruction?

BY DANIEL KORETZ

The educational "reform movement" is many things to many people, but there is no question that stiffened standards and greater accountability are its cornerstone, and achievement testing has been a primary tool for effecting them. Standardized tests, which in the not-so-distant past served primarily as *indicators* of achievement, are now our primary *accountability measures*. That is, tests are increasingly used, not only to show how well students do, but also to judge the competence of the educational enterprise and to hold the participants—students, teachers, principals, superintendents—accountable. Test scores have become the common currency for educators wanting to demonstrate progress to their publics, their superiors, or their skeptical legislatures. As Gordon Ambach has so aptly put it, we have entered a period of "measurement-driven educational policy."

This change had its vocal opponents. A number of educators and educational researchers argued that standardized tests (of the sort now typically used) are inadequate indicators of students' achievement, that they often provide little reliable evidence of the effectiveness of educational systems, and that overreliance on them can distort—even degrade—the curriculum. A few observers also argued that reliance on tests as accountability measures would itself distort scores on the tests, clouding the meaning of score increases.

These objections, however, seemingly did little to slow the juggernaut of test-based accountability. Policymakers seemed confident that, whatever the costs

Daniel Koretz is a senior social scientist with the Education Program of the RAND Corporation and a former classroom teacher. He is the author of Educational Achievement: Explanations and Implications of Recent Trends, published last year by the Congressional Budget Office.



and limitations of test-based accountability, the benefits would be greater. Perhaps the measures are not ideal, the argument went, and perhaps there are some unintended, undesirable side effects, but tests still allow us a reasonable index of the educational system's performance that we need, both to hold schools' feet to the fire and to demonstrate our accomplishments to legislators and the public.

This confidence in test-based accountability, however, faces a new threat, from an unlikely source. What opponents in the educational and research communities were unable to achieve may have been accomplished by a physician in Beaver, West Virginia.

THE LAKE WOBEGON REPORT

Late last year, Friends for Education, an *ad hoc* group headed by Dr. John Jacob Cannell, a general practitioner in Beaver, published a small booklet entitled *Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above the National Average*. The booklet asserts that all states reporting statewide test scores rank above the national average, as do most of the districts from which Dr. Cannell obtained data. These implausible findings earned the booklet the nickname of "the Lake Wobegon report," in reference to Garrison Keillor's mythical community in which "the women are strong, the men are good-looking, and all the children are above average."

Dr. Cannell's inquiry began when he became disturbed by a pattern he observed in his medical practice. Often he would see troubled adolescents whose problems included poor academic performance. But time and time again, Dr. Cannell was assured by school personnel that his patients' achievement, as measured by test scores, was not so bad. Dr. Cannell became suspicious, and his doubts grew sharply when he saw West



ILLUSTRATED BY ROBERT BARKIN

Virginia's statewide results on the Comprehensive Tests of Basic Skills (CTBS), one of the leading commercial norm-referenced achievement tests. The booklet notes:

The most recent scores on the . . . CTBS had West Virginia above the national average at all grade levels tested. Third graders tested at the 65 percentile and sixth graders at the 62 percentile, compared to the national average of 50. . . . West Virginia has the highest percentage of adults without a college education, the second lowest per capita income, and the third lowest college entrance scores (ACT) in the nation. . . .

If West Virginia is above average, what state is below average? How could West Virginia do so poorly on every other indicator of student performance but consistently test above the national average on the . . . CTBS?

Dr. Cannell then requested test score data from all of the states. He found that "thirty-two states test elementary children on a statewide basis, and all thirty-two are testing above the national average." Included among the thirty-two are many states (primarily in the South) that rank, Dr. Cannell notes, particularly poorly in terms of other indicators, such as graduation and literacy rates:

In South Carolina, where per capita income ranks forty-seventh, graduation rate is forty-seventh, and college entrance scores are the lowest in the nation . . . , 62.9 percent of the fourth graders tested above the national average on the CTBS total battery. Although Georgia is behind the nation in per capita income, graduation rate, and Scholastic Aptitude college admission test scores . . . , more than 68 percent of Georgia's second graders tested above the average on ITBS [Iowa Tests of Basic Skills].

In Kentucky, 79.6 percent of the third graders were told that they are above the national average on a modified CTBS test although Kentucky ranks below the nation in per capita income, graduation rate, and college entrance scores And in Tennessee, second graders had a national average group percentile rank of 86 percent in mathematics and language although Tennessee is below average in per capita income, graduation rate, and college entrance scores.

The remaining states did not have, or did not provide, relevant data from statewide testing, so Dr. Cannell obtained scores from the largest districts in each. "All eighteen states had the vast majority of the surveyed districts above the national average," Dr. Cannell reported. "These larger districts often include metropolitan areas with many inner city children, a group often thought to be below the norm." Among Dr. Cannell's examples: Trenton and East Orange, New Jersey; New York City; St. Louis, East St. Louis, and Kansas City, Missouri; and Boston.

These findings led Friends for Education to harsh conclusions about standardized achievement testing:

These tests allow all states to be above average The tests . . . allow 90 percent of the school districts in the United States to be above average. More than 70 percent of the students tested nationwide are told they are performing above average Friends for Education found that these standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press misleading reports on achievement levels.

In other forums, Dr. Cannell has called the tests "deceptive," and his group has filed "consumer fraud complaints" in all 50 states against four publishers of norm-referenced tests.

To the extent that it is accurate, Dr. Cannell's report

In my opinion, there can be no doubt that current norm-referenced tests overstate achievement levels in many schools, districts, and states, often by a large margin.



has even broader implications than he recognized, for it weakens the rationale for test-based accountability as it is currently practiced. The report reveals that one of the benefits of current test-based accountability is scantier than proponents argue. If Dr. Cannell is right, one can no longer say that current accountability-oriented testing at least gives us a reasonable, if crude, view of how well our students are mastering certain skills that we agree are important. Not if every state and most districts are claiming to be above average. That in itself makes the costs—in money, students' and teachers' time, and stress—loom larger in comparison. In addition, there is some evidence, as I will explain, that part of the exaggeration of scores reflects undesirable instructional practices that raise *scores* more than *achievement*.

IS DR. CANNELL CORRECT?

Much of the debate sparked by the Lake Wobegon report has focused on the accuracy of Dr. Cannell's conclusions. Critics have argued that Dr. Cannell does not understand either testing or the mathematics of test scores—a charge that Dr. Cannell freely accepts—and that the report is full of inaccurate and misleading figures.

The Lake Wobegon report does in fact contain many inaccuracies, and many of the figures in it are interpreted incorrectly. Some of these errors are of little practical significance, while others are important. As he notes in his booklet, Dr. Cannell reported the data in whatever form he received it. Districts and states, however, report data in many different ways. In some instances, they report the numbers that Dr. Cannell needed: the percent of students supposedly scoring above the national mean or median. In most cases, however, Dr. Cannell did not obtain those numbers and made do with alternatives: the percentile ranking (in terms of the national distribution of scores) of a district's or state's average or typical student, the jurisdiction's average score in normal-curve equivalents, the percent of students scoring in the fourth stanine or above, or "scores reported in terms of being above average in average percent correct." These alternative measures do not neatly correspond, however, to the percent of children scoring above average.

In some instances, Dr. Cannell's use of the wrong numbers clearly distorts his conclusion substantially. For example, three of the states presented in the Lake Wobegon report as having exceptionally high scores—Hawaii, Indiana, and Nevada—report the percentage of students scoring "in stanine 4 or above." These percentages are presented by Dr. Cannell as yet another indication of the proportion "above average." "Stanine 4 and above," however, does not indicate the proportion of students above the national average; rather, it indicates the proportion of students who are in or above a *broad middle range* of achievement. (The average on a stanine scale is 5, not 4.) On a perfect stanine scale, 77 percent of students would score at stanine 4 or above. The result of this error is to greatly overstate the proportion of students who score above average. In the case of Hawaii, for example, Dr. Cannell also presented the national percentile rankings of the average student, as a parenthetical aside. The latter numbers, which are more closely related to (but not the same as) the proportion

of students scoring above average, are as much as 25 points lower than the percentage of Hawaiian students scoring at or above stanine 4 (and are below 50 in half of the reported instances).

Dr. Cannell's conclusions are also overstated because the report did not recognize the important difference between the average scores of *districts* and the scores of the average *students* in those jurisdictions. Although this may seem counterintuitive, the proportion of districts with average scores above the national average can be quite different from, and is often markedly higher than, the proportion of students in that state scoring above the national average. Dr. Cannell's own figures illustrate this. Quite a number of states reported to him both the percent of students and the percent of districts above the national average. In virtually every instance, the latter number was larger, often by a sizable amount. Thus, even if tests were working as they ought, more than half of all districts could be "above average." The impossibly large proportion of students scoring above average indicates that the figures for districts are inflated also, but less than they seem.

Dr. Cannell's errors are to some extent beside the point, however, for they are not sufficient to call into question his basic conclusion. In my opinion, there can be no doubt that current norm-referenced tests overstate achievement levels in many schools, districts, and states, often by a large margin. To my knowledge, none of Dr. Cannell's critics have disagreed with this judgment. The exaggeration of achievement is particularly pronounced when scores are compared to a hypothetical national norm, but it also often occurs when achievement is compared to an absolute standard. Unfortunately, no one really knows just how serious the exaggeration is, and indeed there is no single answer. The extent of overstatement varies markedly depending on a host of factors: the characteristics of the test used, the recency of the test's norms, the severity of the stakes attached to the test scores (for teachers and administrators as well as students), and so on. Nonetheless, it is clear that the exaggeration is widespread and, in some instances, sizable.

WHAT CAUSES THE LAKE WOBEGON PHENOMENON?

A key question—both for testing policy and for educational practice more generally—is *why* scores are exaggerated. Dr. Cannell blames both test publishers and educators: "Friends for Education suspects that inaccurate initial norms and teaching to the test may be the reason for high scores." Teaching to the test requires the cooperation of teachers, of course, but Dr. Cannell places part of the blame for it on the shoulders of administrators who unduly pressure teachers to raise scores. Furthermore, Dr. Cannell strongly implies deliberate wrongdoing. Consistent with Dr. Cannell's views, much of the ensuing debate has focused on two possible causes: malfeasance and technical weaknesses in testing programs. These are tempting targets. Outright wrongdoing is clearly worth rooting out, and we Americans are always confident that technical problems can be solved.

Both of these causes are important, but the Lake Wobegon problem will not be eliminated unless we

look beyond them to additional factors, less tidy and tractable, that lie uncomfortably close to the heart of today's educational practices and policies. First, a fundamental aspect of the problem rests, not with how tests are constructed, but with how they are currently *used*, as tools of both instruction and policy. Second, the outright wrongdoing that we all want to see expunged fades imperceptibly into a vast, gray area of practices that fall somewhere in between malfeasance and appropriate educational practice. Even when these practices do not represent deliberate wrongdoing, they nonetheless may distort test scores and misdirect or degrade instruction. These aspects of the problem will prove the most difficult to solve, and, in my view, warrant especially intense attention.

TESTING'S NEW ROLE: THE PRESSURE IS ON

In order to understand the roots of the Lake Wobegon problem, one must recognize the tremendous growth in the consequences attached to test scores in recent years. In the not-too-distant past—when most of us were in school—standardized testing was a relatively low-stakes enterprise. For certain students, of course, the consequences of test scores were significant even then—for example, scores were a factor in placement and tracking decisions. But for most students and educators, the consequences were limited. Standardized tests were just one source of information about how well students were doing.

No longer. First, the stakes for students were dramatically raised, as both graduation and promotion between grades were increasingly tied to test scores. This trend began with the spread of minimum-competency tests in the late 1970s, but it has not ended yet. In a particularly extreme case, Georgia recently mandated that a commercial standardized test be used statewide to determine eligibility for promotion from kindergarten to first grade, provoking criticism from early childhood educators and experts in child development.¹ Indiana's new ISTEP testing program, implemented statewide for the first time last March, uses fixed cut-off scores on a customized standardized test to select students for both remedial summer school and retention in grade, beginning in the first grade.² These uses of tests represent a dramatic change in what determines students' progression through school; the judgments of teachers and building administrators, which ideally reflect many diverse indications of students' performance, are replaced in substantial part by students' scores on a single test.*

More recently, the stakes have risen for teachers and administrators too, as test scores have increasingly been used to hold them accountable as well. This shift was in some ways the root cause of the Lake Wobegon phe-

nomenon, for it forced attention to the average scores of groups—schools, districts, and states—rather than just the scores of individual students.* The resulting pressure on educators to raise scores comes from all quarters. The U.S. Department of Education has attempted to use publicity to pressure states to raise scores. States pressure districts and schools in a variety of different ways. Some states have relied on publicity; for example, both Pennsylvania and Kentucky have publicized rankings of schools based on test scores.³ Other states have implemented or plan to implement programs that reward schools with extra funds for good performance on standardized tests. (In most cases, test scores are only one of several criteria that are considered in selecting grantees.)

District superintendents and boards also sometimes turn up the heat. For example, John Murphy, superintendent of the Prince Georges County (Md.) public schools (one of the nation's largest districts), says he keeps charts of the test scores of all of the district's schools on the walls of his conference room—a key component of what the district describes as its “applied anxiety” strategy. Principals sometimes add their own weight to the pressures, and parents weigh in where everyone else has left off.

The amount of pressure finally felt by teachers and administrators varies, but it can be intense indeed and can be more severe than the sanctions embodied in explicit policy statements suggest. One study, for example, found an instance in which high school teachers responsible for preparing students in a mixed middle and high school for a statewide civics examination were told by their principal that they would be assigned part-time to the middle school (apparently considered a punishment) if scores did not rise. Two years later, scores went down marginally, and the two teachers involved were re-assigned to middle school duties. In another instance, the release of figures showing a trivially lower passing rate on a minimum-competency test in one of two middle schools in a district led to a flood of parental requests for transfers out of that school.⁴ These extreme examples may be exceptional (although we don't know that they are), but the pressure they illustrate is not.

SOME SPECIFIC CAUSES OF THE PROBLEM

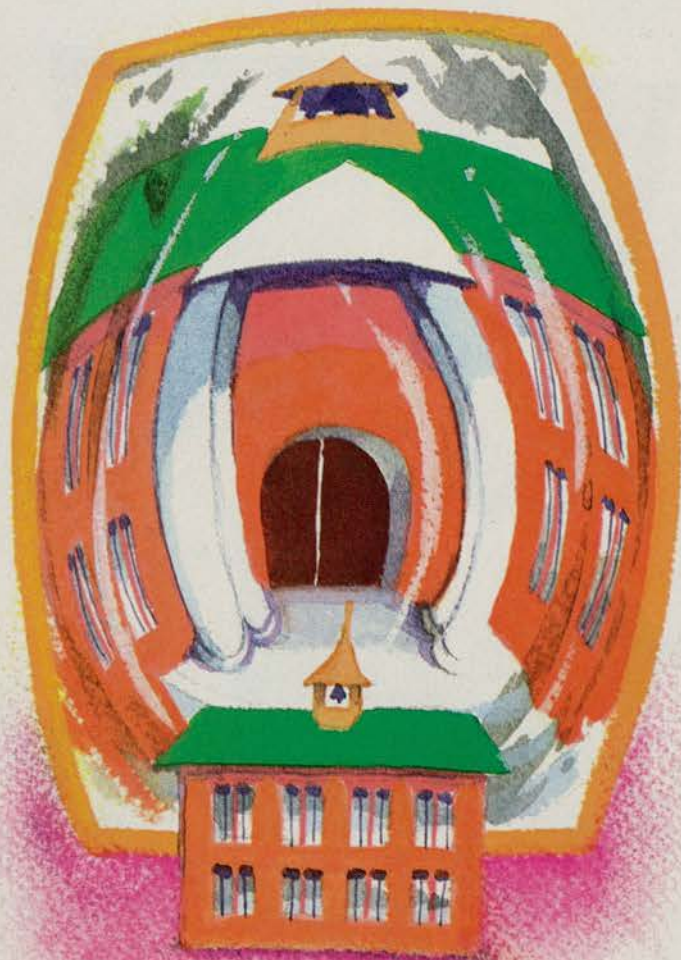
With this context in mind, we can sketch some of the factors that contribute to the Lake Wobegon problem.

Dated Norms. First, it is necessary to describe a bit about how standardized tests are produced. In most instances, publishers sell one edition of a test battery for a considerable period—often, seven years. Before that edition is released, it is administered to a national sample of students to obtain the “norms”—that is, a nationally representative distribution of scores. Scores

*This oversimplifies a bit. For example, in the ISTEP program, a building administrator can seek waivers, on a case-by-case basis, if she believes that a student's failing performance on the ISTEP test is an inaccurate indication of the student's true performance level. As a general rule, though, minimum-competency examinations preempt educators' judgments about promotion and graduation in most instances in which students fall below the established minimum score.

*This emphasis on the scores of groups is a primary cause of the tremendous centralization of authority for testing that has occurred over the past decade. Mandatory statewide testing, only recently the exception, has become the norm, and national testing will be a reality within a few years, as the National Assessment of Educational Progress is expanded to permit state-by-state comparisons.

In another instance, the release of figures showing a trivially lower passing rate on a minimum-competency test in one of two middle schools in a district led to a flood of parental requests for transfers out of that school.



of individual students can then be compared to the national distribution of scores; a given student, for example, may be told that her score falls at the 90th percentile on the national distribution. Students are usually compared to the same set of norms for as long as that edition of the test remains in use. When a replacement edition is readied, the process is repeated, and scores on the new edition are usually "equated" to those on the old edition. In its simplest form, equating involves administering both editions to the same sample of students. This enables the publisher to establish what score on the new form is equivalent to any score on the old edition.

For present purposes, a critical feature of this process is that the norms become increasingly dated until a new edition is introduced. Students (and districts and states) are compared to a national standard that is sometimes more than half a decade out of date.

This causes part of the Lake Wobegon effect. When scores are rising or falling nationwide, comparisons of current scores to these outdated national norms are misleading. Test scores have been going up nationwide for more than a decade in the lower grades and nearly a decade at the senior high school level.⁵ Although a part of this rise is spurious, part is real: students are in fact improving their mastery of the knowledge and skills tapped by standardized tests.* As a result, districts (or states, or whatever) that are merely keeping pace with the national rise in scores will increasingly appear "above average," but only because the "average" that is the standard of comparison is out of date and is therefore lower than the true—but unmeasured—national average at the time the comparison is made. Indeed, even districts that are failing to keep pace with the nation as a whole will become "above average," as long as they are still improving relative to the old national norm.†

This share of the Lake Wobegon phenomenon, then, represents historical happenstance, not malfeasance on the part of the publishers or anyone else. Had Dr. Cannel conducted his investigation during the 1960s and early 1970s, historical trends would have biased districts' and students' apparent performance in the other direction. Test scores nationwide were falling, so the use of dated norms would have made districts that were at or somewhat above the current national average appear below average. Infrequent renorming persisted nonetheless, because it is cheaper than annual norming.* On the other hand, there is no reason why administrators and test publishers could not warn the public that comparisons to old norms are misleading.

Test Selection. A second and less tractable cause of the Lake Wobegon phenomenon lies in districts' latitude in selecting tests. Although the major tests correlate highly

*This need not imply, however, that students are improving in terms of skills *not* tapped by the tests—for example, some of the higher-order skills that multiple-choice tests do not assess well.

†This aspect of the problem, one should note, only affects comparisons of the scores of a student or jurisdiction to national norms. Estimates of their achievement relative to some absolute standard would be unaffected by a failure to take national trends into account.

with each other, there are important differences among them in content, emphasis, and format. In selecting tests, administrators will typically look for the ones that come closest to matching their curricular objectives. (Within limits, that is of course a reasonable path to follow. To gauge the effectiveness of instruction, one would not want to use a test that contains a lot of material that students have not been taught. Given the history of court decisions about testing, it is also a sensible legal precaution.) Moreover, once a test is in place, administrators can modify the curriculum to align it even more closely with the test—an aspect of the general problem, discussed below, of teaching to the test.

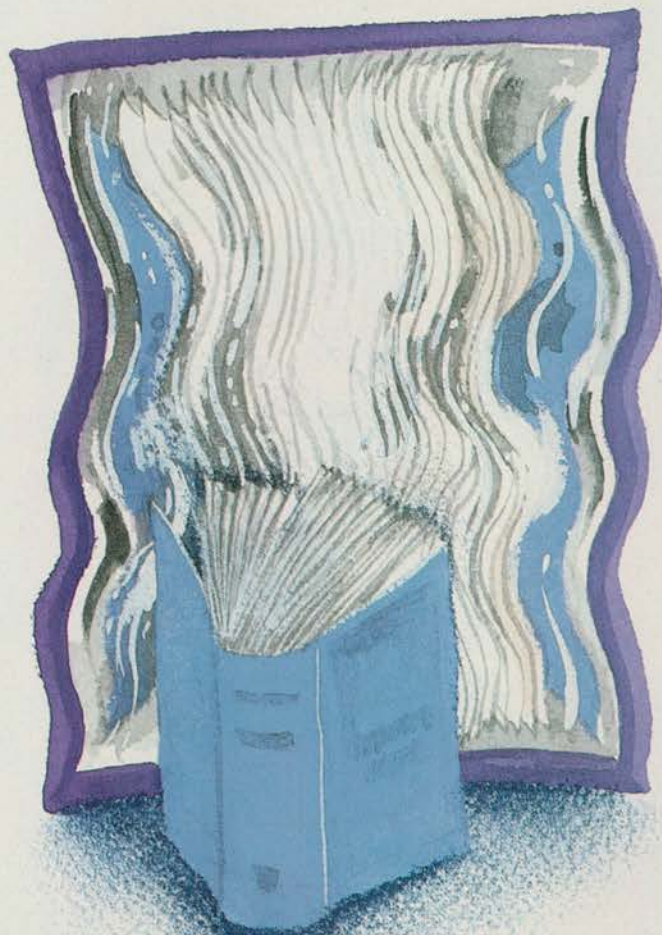
These decisions may make good educational and legal sense, but they nonetheless can inflate districts' and states' apparent performance relative to the national norms. Ideally, the norming sample is representative of the nation as a whole, not of the subset of districts whose curricula most closely match the content of a given test. If administrators do a good job of choosing tests that match their curricula, their students will tend to score higher on the test than the norming sample did, even if their level of achievement is in some broader sense equivalent, simply because their curricula match the test more closely and thus prepare them better for it. Although no one can estimate how much this factor contributed to the Lake Wobegon effect nationwide, we do have clear evidence that it sometimes makes a large difference.⁶ This factor is a likely culprit when jurisdictions one would not expect to do well have high scores in the first year of administering a recently revised test—several instances of which were emphasized in the Lake Wobegon report.*

The leeway districts have in choosing tests also raises the question of how scores ought to be reported to the public. As in the case of out-of-date norms, some administrators and all test publishers realize that this factor sometimes makes jurisdictions appear more successful than they really are. At the same time, they typically cannot know the precise magnitude of the resulting bias, so they would be unable to adjust the scores even if they wanted to. There is no straightforward solution to this dilemma, but accompanying the release of scores with a warning to this effect would be more candid and informative than simply reporting scores without qualification.

The desire to select a test that is aligned with the curriculum takes its most extreme form when districts or states arrange with publishers to provide them with

*This latitude in selecting tests also has less responsible manifestations. In some cases, administrators have prolonged the use of out-of-date norms to make their districts look better than they ought. In other cases, a district's likely performance on a test, rather than congruence with broader curricular goals, has been an explicit criterion for selecting tests. One test developer reported to me an incident in which a superintendent of a high-scoring district was searching for a testing program that would make his district score less well, in order to increase the number of students eligible for Chapter 1 compensatory-education funding. It is important to bear in mind, however, that even if these inappropriate choices were utterly eliminated, the appropriate leeway jurisdictions have in selecting tests that correspond to their instructional goals will often result in misleading comparisons to national norms.

The desire to select a test that is aligned with the curriculum takes its most extreme form when districts or states arrange with publishers to provide them with "content-customized" versions of standardized achievement tests.



"content-customized" versions of standardized achievement tests. The state or district requests that the publisher alter an existing test to more closely match the jurisdiction's curriculum. This in itself is not problematic; the difficulty arises when scores on the customized test are expressed in terms of the publisher's national norms—which reflect performance on the original, not the customized, test.

The appropriateness of reporting scores on customized tests in terms of norms from the parent test is a technically complex issue that cannot be adequately addressed here. It is nonetheless clear that this process can contribute to the Lake Wobegon phenomenon. Linking a customized test to norms from the parent test requires that the relative difficulty of the two tests be appraised. But the difficulty of a test item is not an inherent trait of that item; rather, the success rate of students on any given item depends on their exposure to the relevant content. If teachers shift emphasis toward items that are added in the customizing, those items will become easier. Conversely, if other items on the test receive less emphasis, they will become harder than they were for comparable students in the norming sample. But if the test was made longer by customizing, these latter items will count less; if some were dropped to keep the test's length constant, then they obviously will not count at all. Thus the more teachers tailor their teaching to a customized test, the better scores will seem in terms of the norms on the parent tests.*

Which Students Are Tested. Another likely contributor to the Lake Wobegon phenomenon is the fact that the actual administration of tests is in the hands of the districts that purchase them. This has several ramifications. Perhaps most important, there need be no consistency in decisions about who is tested. Which students should be considered sufficiently handicapped, for example, or sufficiently limited in English-language proficiency, that they should be excluded from testing? What effort should be made to test truants, or disruptive students? Because many of the students about whom such decisions must be made will score well below the average, these decisions can have a major impact on the average score in a school, district, or state. I am not aware of any systematic evidence about trends in these selection decisions, but it is probably reasonable to suspect that the current pressure to raise test scores has led to changes in some districts that have inflated the rise in scores.

Non-secure Tests. In addition, because the districts administer the tests, they cannot be considered "secure." The tests that will be given next year are physically present in the district; moreover, after a number of years, teachers will presumably have seen many of the items even if the tests are kept locked up, because of their experiences as proctors. This is one of several factors that facilitate teaching to the test.

*Some educational researchers argue that even without this shift in instructional emphasis, customized tests that are not representative of the parent test will often give misleading estimates of performance, depending in part on the methods used to equate the customized and parent tests. The point made here, however, does not hinge on this argument's being correct; it depends only on differences in instructional emphasis.

Talk of non-secure tests immediately raises the specter of cheating—for example, providing students with items from the test. We know that cheating occurs. In one recent case, for example, a large, high-scoring, suburban district disciplined several teachers for providing students with a computer program that drilled them on items from an upcoming achievement test. A few years earlier, the state education agency in that same state noticed implausibly high scores from several other districts; teaching actual items was again the culprit.

Cheating is not limited to teachers. In one mid-Atlantic state, for example, a district's mathematics supervisor discovered that the state's minimum competency test used shaded figures for problems involving the calculation of areas and unshaded figures for perimeters. She then told a number of teachers (primarily, special education teachers) that they could give their students a simple rule: multiply if shaded, add if not.⁷ In another instance, actions that many observers would consider cheating occurred at the level of district administration. Nearly a decade ago, San Diego was confronted with a court order demanding that test scores be raised in certain "minority-isolated" schools in lieu of desegregation. The court specified the dates by which given proportions of students in those schools were to exceed the national median score on the CTBS. High-level district staff responded by writing their own mastery-learning curriculum, portions of which incorporated content from the CTBS.⁸

No one knows how much of this outright cheating has occurred, or what share of the Lake Wobegon phenomenon can be attributed to it. No one has systematically collected information on the instances that have been identified, and, in a system of over 15,000 local districts and over two million teachers, a considerable proportion probably goes unnoticed or unreported. Still, while the current escalating pressure to raise scores has no doubt increased the temptation to cheat, I would suspect that even the total elimination of outright cheating would make only a very modest dent in the Lake Wobegon phenomenon.

TEACHING TO THE TEST

I suspect that outright cheating is what Dr. Cannell had in mind when he fingered "teaching to the test" as a cause of the Lake Wobegon phenomenon, but teaching to the test is in fact a much larger, vastly more important, and much less tractable problem than frank cheating. Indeed, teaching to the test is one of the most pressing issues of educational policy and practice today. For teaching to the test can do more than inflate test scores; it can also degrade instruction and impede learning. Moreover, teaching to the test is a problem that can involve the entire educational system. While it ultimately gets played out in the classroom, its origins are often found elsewhere—for example, in decisions made and directives issued by administrators, school board members, and legislators.

But what actually constitutes "teaching to the test?" Here that gray area between appropriate practice and malfeasance is particularly large and especially hard to delineate. Teaching to the test is generally a derogatory

(Continued on page 46)

THE LOSS OF JOBS AND THE RISE OF THE UNDERCLASS

A Review of "The Truly Disadvantaged."

BY NORMAN HILL

The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy by William Julius Wilson (Chicago: the University of Chicago Press, 1987, 254 pp., \$19.95.)

In the 1960s, the civil rights revolution took hold: legal segregation was ended, new educational and job opportunities opened up for blacks, affirmative action programs were launched. Two decades later, large numbers of blacks are firmly in the middle class, college attendance is way up, the wage gap has declined dramatically, the number of blacks holding political office has grown rapidly, and many blacks are moving up in their companies' ranks.

But the revolution was never completed. In the ghetto, things have never been worse. In the wake of an extraordinary revolution in civil rights, the nation's poorest blacks have been left far, far behind.

What led the underclass to burgeon and what policies might best rescue its members from a life of misery and poverty are among the most pressing and oft-discussed questions of social policy today.

THE STATISTICS are harrowing: Daniel Patrick Moynihan, in his 1965 report on the disintegration of the black family, expressed concern that one-quarter of all black families were headed by women. Twenty

Norman Hill is the president of the A. Philip Randolph Institute, a national organization of black trade unionists committed to increasing black involvement in trade unions and in political life.

years later, that figure had jumped to an alarming 43 percent. In 1959, only 15 percent of black births were out of wedlock; by 1982, it was 57 percent. In 1960, 42 percent of children born to black teenagers were illegitimate; by 1983 it was 89 percent. Finally, the poverty rate for black female-headed families in 1982 was a staggering 56.2 percent.

Figures regarding education are equally disturbing. In Chicago, of 25,500 ninth grade black students enrolled in segregated, nonselective high schools who made up the class of 1984, 16,000 did not graduate.

And although blacks constitute 13 percent of the urban population, they account for over 50 percent of all city arrests for violent crime. In 1984, the rate of black imprisonment was 6.25 times greater than the rate of white imprisonment. Overall, 46 percent of the country's prison population is black.

The terrible pathologies of the inner-city ghettos—illegitimacy, family dissolution, welfare dependency, soaring dropout rates, teenage pregnancy, alcohol and drug abuse, crime—have, over the last two decades, been attributed to a wide range of causes. In fact, no other social problem has generated as much anguish, debate, and controversy among politicians, civil rights activists, and social scientists trying to determine its roots and possible policy solutions. Perhaps inevitably, analysis of sociological data has often split along ideological lines. After Moynihan and his report were sharply criticized as racist by many in the civil rights community, many liberals—fearful of being branded racists or of causing a rift in the civil rights coalition—either ignored the issue of the underclass or generally blamed its rise on factors for which blacks themselves clearly could not be blamed, such as the lingering miasma of slavery, the persistence of racism, and the



ILLUSTRATED BY SALVADOR BRU

allegedly inevitable result of a massive migration to the industrial north of southern black sharecroppers unused to the incentives and schedules of factory work. The liberals' unwillingness to seriously analyze the problem, coupled with the apparent failure of liberal social policies enacted during the Great Society to alleviate the plight of the underclass, emboldened conservatives to challenge liberal assumptions as to the cause and solutions to the problem. Conservatives, most notably Charles Murray in his provocative book, *Losing Ground*, have argued that federal anti-poverty programs and the current welfare system reward childbearing, single motherhood, and joblessness, and have engendered a permanent "culture of poverty" informed by so-called ghetto values that appear to defy conventional socioeconomic solutions.

IN AN important and compelling new book, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, William Julius Wilson, professor of sociology at the University of Chicago, debunks many of the liberal and conservative analyses regarding the underclass and offers a perceptive, persuasive argument as to its evolution and character. *Simply put, Professor Wilson places most of the blame for the precipitous rise in the black urban underclass squarely on the rapid, systemic, and structural changes in the economy in the 1970s, which left black unemployment at a spectacularly high rate.* While acknowledging the persistence of racism as a factor hindering black economic progress, he puts the overwhelming blame for the rise of the underclass on the loss of millions of low- and semi-skilled jobs in America's large urban manufacturing centers and the steep increase of joblessness among black

males, a trend that he says has destabilized the black family and, by extension, the black community. In so doing, Wilson breaks through the stultifying old arguments and debates and sets a new public policy agenda for the civil rights movement and the nation. Before detailing his findings and proposals, let me review the old arguments and their inadequacies as illuminated by Wilson.

First, he challenges the oft-repeated assumption that slavery, which often led to the splitting up of black family units, has had a long-term, continuing impact on black family cohesion. Wilson cites studies that indicate that, contrary to this view, post-slavery blacks defied great odds and successfully pulled their families together: In the first quarter of this century, between 70 and 90 percent of black households were "male present" and a majority were nuclear families. Moreover, in 1940, under 18 percent of black families were headed by women, most of whom were widows. Today, that figure approaches 43 percent, and the overwhelming majority of women heading families have *never* been married. Citing these statistics in a recent article, author Pete Hamill wrote: "How could slavery have a greater corrosive effect on the black family *today*, almost half a century later, than it had in 1940? The question contains its own answer; it couldn't."

Yet another theory challenged by Wilson is the notion that a segment of the 6.5 million blacks migrating from the rural south to the northern cities between World War I and the mid-1960s brought with them a "sharecropper culture" characterized by an ethic of dependency, large out-of-wedlock families, and poor education. The baggage of these cultural factors, when brought north by the migrants, according to Nicholas

Lemann, writing in *The Atlantic Monthly*, had a debilitating effect on a segment of the black population and inhibited social and economic advancement and mobility out of the ghetto. But Wilson found that studies on urban poverty and recent migration consistently show that "southern-born blacks who have migrated to the urban north experience greater economic success in terms of employment rates, earnings, and welfare dependency than do those urban blacks who were born in the north. One study found that northern-born blacks were more likely to receive welfare than southern-born blacks, despite the fact that the level of education of blacks born in the north was higher than that of blacks born in the south. In addition, the migrants (both men and women) have tended to have higher labor-force participation rates and lower unemployment rates than black natives in the cities in question."

"If contemporary discrimination is the main culprit, why did it produce the most severe problems of urban social dislocation during the 1970s?"

But perhaps the most commonly offered liberal explanation disputed by Wilson is that the social deterioration in the urban ghetto is the result *primarily* of *continued* discrimination and racism. (Clearly, *historical* bias has had a devastating effect.) Noting that the pathologies in the inner cities "did not reach catastrophic proportions until the mid-1970s," Wilson asks: "If contemporary discrimination is the main culprit, why did it produce the most severe problems of urban social dislocation during the 1970s, a decade that followed an unprecedented period of civil rights legislation and ushered in the affirmative action programs?"

According to Wilson, the race-specific explanations are even more problematic in view of the dramatic economic progress of the black middle class during that same period, a phenomenon Wilson explored in his earlier book, *The Declining Significance of Race*.

Having effectively dispensed with the arguments often put forward by liberals, Wilson turns to the conservatives, using the bulk of his book to debunk their hypotheses, which tie the growth of welfare and anti-poverty programs to the rise of the ghetto underclass. Referring to Charles Murray's central thesis that federal programs have contributed to the rise of the underclass and poverty by changing the rewards and penalties that govern human behavior, Wilson points out that the real value of the two programmatic mainstays of welfare—food stamps and Aid to Families With Dependent Chil-

dren (AFDC)—have declined dramatically since 1972 because the states have not made adjustments for inflation. And while benefit levels have fallen in real terms over the last 10 years, illegitimacy ratios have continued to rise. In addition, in 1975, Congress enacted the Earned Income Tax Credit, which further expanded the comparative advantages of work over welfare for the poor.

"Thus, if welfare incentives lead to black joblessness and family dissolution, as Murray argues, these trends should have reversed themselves in the 1970s, when the relative advantage of work over welfare increased sharply."

Wilson further challenges Murray's contention that there was more job growth in the 1970s, when poverty rates went up, than in the 1950s, when the poverty rate dropped, noting that recent job growth has been insufficient to handle the unusually large numbers of women and babyboomers entering the job market, resulting in an increase in unemployment.

IN WILSON'S view, the most salient factor contributing to the social breakdown in inner-city ghettos is the rapid evaporation of job opportunities for black males as a result of structural changes in the economy. The shift from goods-producing to service-producing industries, the increasing polarization of the labor market into low-wage and high-wage sectors, technological innovations, and the relocation of manufacturing industries out of the central cities have all had a disproportionate impact on black males.

In the last two decades, for example, the four largest northern industrial cities have lost millions of low- and semi-skilled manufacturing jobs, no longer allowing the uneducated or untrained to sell the only commodity they have—muscle power. "Unfortunately," notes

LOSS OF BLUE-COLLAR JOBS
FROM THE CENTRAL CITY

Central-City Jobs in Industries, by Mean Education of Employees, 1970 and 1984 (figures in thousands)

City and Educational Mean of Industry	Number of Jobs		Change 1970-84
	1970	1984	
New York			
Less than high school	1,445	953	- 492
Some higher education	1,002	1,241	239
Philadelphia			
Less than high school	396	224	- 172
Some higher education	205	244	39
Boston			
Less than high school	168	124	- 44
Some higher education	185	252	67
Baltimore			
Less than high school	187	114	- 73
Some higher education	90	105	15
St. Louis			
Less than high school	197	108	- 89
Some higher education	98	96	- 2

Source: John D. Kasarda, "The Regional and Urban Redistribution of People and Jobs in the U.S.," draft paper prepared for the Committee on National Urban Policy, National Academy of Sciences, Washington, D.C., 1986.

In Wilson's view, the ghetto's social chaos stems, in the first instance, not from a lack of a particular set of values, but, quite simply, from the lack of jobs.

Wilson, "essentially all of the national growth in entry-level and other low education requisite jobs have accrued in the suburbs, exurbs, and nonmetropolitan areas far removed from growing concentrations of poorly educated urban minorities."

At the same time, job growth in the urban areas was primarily in fields requiring higher education levels, which "created a serious mismatch between the current educational attainment of minority residents in large northern cities and the changing education requirements of their rapidly transforming industrial bases." This mismatch has contributed to "both unemployment rates and labor force dropout rates among central-city blacks [that] are much higher than those of other central-city residents, and [explain] why black unemployment rates have not responded well to economic recovery in many northern cities." From 1960 to 1984, civilian labor force participation* rates among blacks aged 20 to 24 plummeted from 90.4 to 77.2 percent; among those aged 18 to 19 it fell further—from 71.2 to 55.4 percent. And among blacks living in central-city ghettos, the increase in unemployment has been of monumental proportions. Wilson shows that in Chicago, in 1970, unemployment surpassed 15 percent in just *one* area of the city (i.e., in one census tract); by 1980, unemployment was over 15 percent in 15 areas—all heavily black—and over 20 percent in 10 of those areas.

This dramatic increase in joblessness, particularly in the younger age groups, has led to a shortage of employed "marriageable" men capable of adequately supporting a family. Noting that most men and women marry within their race and near their age, Wilson provides an index of the "male marriageable pool." According to the index, the number of marriageable black men aged 16-44 in northeastern and north central states (where the growth of the underclass has been most pronounced) between 1960 and 1980 dropped by just over 12 men per 100 black women, leaving 50 marriageable black men per 100 black women. (During this period, the male marriageable index for white women in the same regions remained basically constant, at 80 to 90 marriageable white men per 100 white women.)

*This rate is actually based on census figures for "black and other races"; but since 90 percent of this population is black, the statistic is a general reflection of the black population.

Wilson concludes that the increasing rate of joblessness among black men is "a major underlying factor in the rise of black single mothers and female-headed households." In other words, says Wilson, it is the domino effect of the rapid decline in manufacturing and low-skilled jobs—leading first to a lack of employed and "marriageable" men, and then to a rise in single-parent, female-headed households—that has brought about the range of social pathologies in the inner-city ghettos: crime, drug addiction, illegitimacy, welfare dependency, etc.

Simultaneous with the great increase in joblessness, there has been a dramatic flight of the middle and working classes from ghetto neighborhoods. Before the civil rights movement—and greater open housing and job opportunities—these socially stable groups lived in the ghetto and acted as "an important social buffer that could deflect the full impact of the kind of prolonged and increasing joblessness that plagued inner-city neighborhoods in the 1970's and early 1980's. . . ."

IN RELATING his analysis, Wilson acknowledges his debt to the late civil rights activist Bayard Rustin, who predicted over 20 years ago that the social mobility unleashed by civil rights legislation would drain the ghettos of the entrepreneurs, businessmen, professionals, educators, and churchmen who could more readily integrate into mainstream society. The absence from today's ghetto of viable businesses, cultural life, stable families, and working men and women who can serve as community anchors and leaders of community organizations has had a devastating impact on those forced to remain in the inner cities, particularly the young. The underclass is now socially isolated. "Thus, in such neighborhoods the chances are overwhelming that children will seldom interact on a sustained basis with people who are employed or with families that have a steady bread winner."

This analysis leads Wilson to reject conservative analyses that focus on dismantling the "culture of poverty." He prefers "to emphasize the concept [of] social isolation." This "does not mean that cultural traits are irrelevant in understanding behavior in highly concentrated [poverty] areas; rather it highlights the fact that culture is a response to social-structural constraints and opportunities." In Wilson's view, the ghetto's social chaos stems, in the first instance, not from a lack of a particular set of values, but, quite simply, from the lack of jobs. And in their need for jobs, Wilson points out, the black poor have more in common with the white poor than with middle- and upper-class blacks.

It's in this context that Wilson resists, as a sufficient response, race-specific programs to address the plight of the black underclass. As have numerous others, Wilson concludes that affirmative action programs most benefit more advantaged blacks: those with higher incomes, greater education and training, and more prestigious occupations. After all, affirmative action applies only to jobs that already exist and helps only those educated or skilled enough to fill them.

Wilson also rejects piecemeal programs directed solely at poor minorities, arguing that such programs
(Continued on page 45)

AMERICA'S TEXTBOOK FIASCO

A Conspiracy of Good Intentions

BY HARRIET TYSON-BERNSTEIN

"If the customer wants a pink stretch Cadillac, I may think it's tacky and wasteful, but I would be a fool to produce a fuel-efficient black compact if nobody is going to buy it."

—textbook executive—

"The most frustrating part of the job is that the user isn't the buyer."

—textbook author—

"The books are all alike anyway, so we don't even bother to read them. We go for the publisher who gives us the biggest freebie package."

—curriculum supervisor—

"They have broken up learning into bits no larger than an eyelash, and the kids aren't able to sweep up the pieces."

—textbook editor—

"All I know is we can make those publishers do what we tell them to do. They support our curriculum, and our scores are going up."

—city superintendent—

IMAGINE a public policy system that is perfectly designed to produce textbooks that confuse, mislead, and profoundly bore students, while at the same time making all of the adults involved in the process

Harriet Tyson-Bernstein is a freelance writer, researcher, and consultant on textbooks, teacher evaluation, and other education issues. She has served as president of the Montgomery County (Md.) Board of Education and director of the project on textbook reform sponsored by the Council of Chief State School Officers and the National Association of State Boards of Education. This article is adapted from The Textbook Fiasco: A Conspiracy of Good Intentions, a report published by the Council for Basic Education, an organization that promotes a rigorous liberal arts education for all students. AFT members may order copies of the full report in book form at the discount price of \$7 each plus \$3 postage and handling from: CBE, 725 15th St., NW, Washington, DC 20005.

look good, not only in their own eyes, but in the eyes of others. Although there are some good textbooks on the market, publishers and editors are virtually compelled by public policies and practices to create textbooks that confuse students with non sequiturs, that mislead them with misinformation, and that profoundly bore them with pointlessly arid writing.

None of the adults in this very complex system intends this outcome. To the contrary, each of them wants to produce good effects, and each public policy, regulation, or conventional practice was intended to make some improvement or prevent some abuse. But the cumulative effects of well-intentioned and seemingly reasonable state and local regulation are textbooks that squander the intellectual capital of our youth.

Some critics of the American textbook system would
(Continued on page 23)

ILLUSTRATED BY DAN SHIBBO



MUDDLE BY MANDATE: A FICTIONALIZED ACCOUNT

"I saw our whole program get shoved to the back of the table because our textbook box wasn't as pretty as our competitors'."

—sales director—

AN EFFECTIVE reform effort depends on a full understanding of the economic and political realities of how textbooks are written, published, and adopted. Before going on to our fictionalized account of a textbook's journey through the labyrinthine adoption system, we pause to explain a few basic facts.

The first and most influential reality is statewide adoption. The 22 adoption states, most of them in the south and west, usually require local districts to select from books on an official state list, even if the local district is expected to pay for them. In some states, a district may obtain a waiver or buy unapproved books out of local funds. If a publisher can get his textbook on the state-approved list in an adoption state, he has a far better chance of selling his books in that state because they will qualify for state reimbursement to local districts that buy them.

Another major variable in the economics of textbooks emerges from two factors. One is the timing of major adoption contests. Adoption state codes specify the length of time an adopted book will be used (the adoption cycle) and the years in which adoptions of new books will take place. Sometimes the adoption cycles of major states for particular subjects coincide, but often they do not. California, for example, may stage a math adoption in 1987 and Texas in 1989.

The effects of staggered adoption cycles on textbooks must be understood in the context of a second factor, the nearly universal insistence on a recent publication date. Thus, if a publisher prepares a math series for California in 1987, Texas may shun his book in 1989 because it is "out of date." In order to sell in as many markets as possible, publishers are increasingly driven to produce new editions frequently. New editions are now issued about every two years. Naturally, textbook publishers prefer to stage their new editions to coincide with the adoption cycles of the more



lucrative markets, like California and Texas.

New editions, though, are very costly since the manufacturing costs—not the editorial costs—are the larger share of expense. Publishers can legally qualify for a new publication date by changing only a small percentage of pages in the book, so they generally avoid extensive revisions in order to remain cost competitive. The changes are more often cosmetic than substantive. Black and white illustrations will be upgraded to four-color plates. New pedagogical buzz words will replace old ones. But errors and anachronisms may go uncorrected. The market's blind passion for newness contributes enormously to the escalating cost of textbooks without achieving its ostensible objective—up-to-date, accurate content.

Defenders of the statewide adoption system claim that large-volume purchases save them money. Although there is evidence that prices are slightly lower in adoption states, the savings are trivial when one considers the extravagance encouraged by law and regulation. State and local business agents may haggle over pennies or try to extract free textbooks or complimentary extras from publishers, but they preside over policies and practices that make textbooks expensive without necessarily making them good.

Foremost among these costly policies is the demand for new editions; but the practice of giving the contract to the publishers with the largest package of "free" add-ons runs a close second; and the practice of rewarding publishers who produce the largest array of extras (audio and video cassettes, Spanish-language editions, laminated maps, globes, etc.), even though the districts don't use them, is another major contributor to the cost of textbooks.

Fly-by-night selection procedures reward publishers who emphasize expensive packaging, covers, and graphics. Extensive content requirements make the books hefty, and therefore expensive as well.

With these fundamental dynamics in mind, we now turn to a fictionalized account of the development and adoption of a textbook. Our purpose is to show how various actors in the drama struggle with their own career demands, how they cope with absurd time constraints, and how they unknowingly contribute to the further fracturing of a structurally fractured process.

Through this fictitious account, we protect the identities of willful states, misguided experts, cunning marketeers, and overworked teachers and administrators, and try to give the reader an understanding of how textbooks have come to be the way they are despite everybody's best intentions. The story is fictional but not false. Although the events in the story could not have happened in any one state or publishing house, every seemingly outlandish event has occurred somewhere in the recent past. And because textbook publishers aim at a national market, the state and local behaviors depicted here have had an impact everywhere. Admittedly a caricature and definitely playful in tone, our story is designed to give the reader insight into the way things really work.

WITH ONLY 14 months to go before the next textbook adoption in the sovereign State of Nirvana, Dr. Harvey, an official in the state education agency, begins to assemble a task force of educators to draft a new set of curricular goals and objectives for reading, science, and history. The task force will produce documents that specify not only what teachers around the state should be teaching, but also what material the textbooks should cover—and in what order.

Dr. Harvey is well aware of pressures from the state legislature and the chamber of commerce to bring student test scores in Nirvana up to the national average. With so many other states attempting the same thing, the national average score is bound to rise. So Harvey wants to engineer a rise in student achievement quickly while that average is still fairly low. The best way to do that, he thinks, is to ensure a tight correspondence between the curriculum, the test, and the textbook.

Harvey invites Dr. Helmut, the state agency's testing expert, to join the curriculum/textbook adoption task force. Helmut's background and expertise will help keep the members on task—that of raising test scores.

Both Harvey and Helmut believe that students should not be tested on material they have not had the chance to learn. Indeed, recent lawsuits have challenged test-based promotion and graduation policies on the grounds that students had not been taught the material that appeared on the test. So for both ethical and legal reasons, Harvey and Helmut are determined to align all of the elements of the program with great precision.

Harvey also invites Professor Hamilton, a leading light in the College of Education at Nirvana State University, to contribute his expertise to the task force. Hamilton recently published a paper, "A Multivariate Analysis of Canonical Word Frequencies in Junior High School Textbooks: Discourse Processing in Relation to Substance Abuse as a Special Case," and is very excited about the implications of his research for curriculum reform and textbook design. Harvey also appoints another professor, Dr. Humble, from Faraway State Teachers College, to forestall criticism rising out of the rivalry between the two campuses. Dr. Henry, Superin-

(Continued on page 40)



(Continued from page 20)

disagree with our "good intentions" hypothesis. It could be argued that adoption policies that fail to secure serious reviewers, fail to give them adequate time to review materials, and fail to compensate them are not well intended. The case could be made that special-interest groups that fight for their own particular views to the exclusion of others are not well intended. Publishers who concern themselves only with sales could be accused of bad intentions. But more likely, most school boards, and perhaps even top administrators, are simply unaware of the relationship between superficial adoption policies and superficial textbooks. Zealots and idealogues, however distasteful they may be to their opponents, generally believe they are protecting children from harm. And publishers, as businessmen, cannot be expected to be idealists; their first obligation is to return a profit, not to render a public service.

You might suppose that such a system could not exist. If it did, wouldn't those responsible for the educational system change it? Or if not, you might suppose that citizens would rise up in sufficient force to create a better way. But you would be mistaken; such a system does exist, and there is at the moment no noticeable effort to change it.

WHAT IS A GOOD TEXTBOOK?

The very first textbook given to young children in school ought to be so delightful that they want to read more books. Children would not only be thrilled by their growing ability to crack the code of written language, they would also be ushered into the riches of our culture. The skills taught in reading books would be closely related to, and not divorced from, the content of the stories. Mathematics books for young children would help them think mathematically and let them practice their skills on problems worth solving. *All* books written for elementary students ought to contain information that is important to adults as well as children. Parents, as well as students, would enjoy reading them.

Books for older students ought to have a theme or

purpose that is crystal clear. Topics and facts would support, and not distract from, the overall theme. Important and difficult topics—gravity, for example, or the constitutional system of checks and balances—would be presented with enough depth for students to understand. Information about the lives and cultures of minorities, women, workers, or ordinary people would not be stuck on gaudily, but integrated into the text. Controversy—so essential to both democracy and intellectual growth—would be embraced rather than avoided. Students would share in defeats as well as triumphs of those who shaped history or built bodies of knowledge.

Facts ought to be accurate. Questions and exercises ought to encourage students to think rather than force them to hunt down trivial details. Chapter summaries would forge essential connections between ideas; they would not merely be cheat sheets for the questions at the chapter's end.

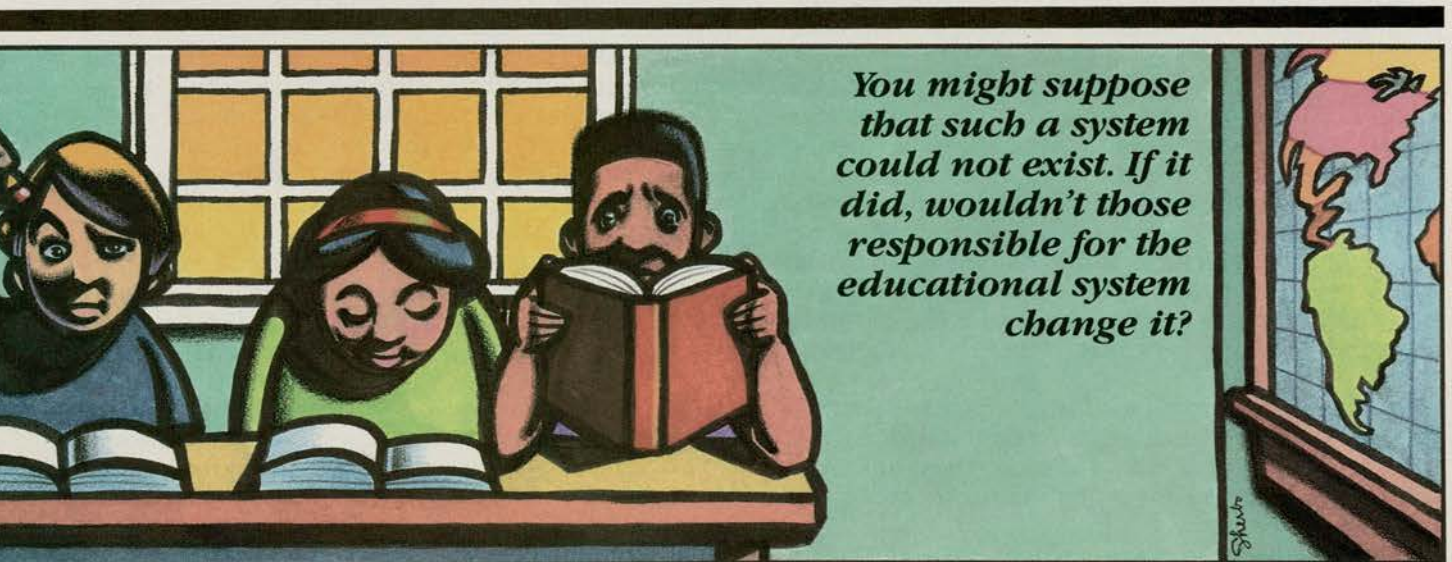
Good textbooks for any grade level or subject ought to be written so students can benefit from the book independently. If students miss a lecture, are absent from school, or merely want to review what was covered in class, they would read the book with some prospect of gathering meaning. The author ought to give students reason for persisting through inevitable patches of drudgery. Most important, the book would be written so students can remember what they read.

Sadly, very few of today's textbooks meet such criteria. Although the flaws in today's textbooks vary according to grade level and subject, two serious flaws afflict the vast majority of commercially prepared materials for schoolchildren: writing is poor, and books treat most topics so superficially that students can't make sense of what they are reading.

THE BAD WRITING PROBLEM

Rabbit said: "I can run. I can run fast. You can't run fast."
Turtle said, "Look Rabbit. See the park. You and I will run.
We'll run to the park."

Rabbit said, "I want to stop. I'll stop here. I can run, but



You might suppose that such a system could not exist. If it did, wouldn't those responsible for the educational system change it?

turtle can't. I can get to the park fast." Turtle said, "I can't run fast. But I will not stop. Rabbit can't see me. I'll get to the park."

In this nearly unrecognizable version of "The Hare and the Tortoise," the main points of the fable—the danger of cockiness and the value of persistence—are utterly lost. Its repetitions are pointless at best and boring at worst. Unfortunately, this kind of editorial mangling is typical of material our children are forced to read in the early grades of school.

Susan Ohanian, in "Ruffles and Flourishes," (*Atlantic Monthly*, September 1987) says, "Basal readers can be criticized on lots of grounds. Their worst fault, I think, is that for no good reason, they squeeze the juice out of some very fine tales." Comparing a passage from the Paul Leysac translation of Hans Christian Anderson's "The Emperor's New Clothes," she finds the following:

"Magnificent!" "Excellent!" "Prodigious!" went from mouth to mouth, and everyone was exceedingly pleased.

had been changed to:

"How marvelous," they echoed the emperor. "How beautiful."

And in "How the Camel Got His Hump," Ohanian notices that Kipling's "a great big lolling humph" has been changed to "a great big humph." "You lose a great big lolling lot when you lose the humph's gerundive," she writes.

Beatrix Potter, whose judgment about what children can and will read is substantiated by the choices made by children at libraries and parents at bookstores, said that children "like a fine word occasionally." Researchers have also demonstrated that kids, as well as adults, like an occasional big word that is delicious to say.

Not only have the fine words been taken out. The basic elements of good storytelling often get lost in the editing process. In his study of elementary readers, Bertram Bruce of Bolt, Beranek and Newman, Inc., found that the conflict essential to any good story has often been suppressed. Moreover, the engaging voice of a narrator, which can bind readers to the story and give them glimpses into characters' thoughts, is often missing.

Comparing the original Judy Blume story "Freckle Juice" with the textbook version, Bruce shows what happens to an otherwise-good story about a freckled boy who thinks his life would be happier if he could get rid of his freckles. In the story, a friend sells the boy some "freckle juice," a vile concoction to smear on his face. In the textbook version, the reader gets little information about why the boy doesn't like his freckles or why he allows himself to be gulled by his friend. "Without these elements," says Bruce, "the story makes little sense."

In general, the writing in elementary schoolbooks is choppy, stilted, and monotonous. Worse still, the words and the phrases that help a novice reader infer the correct relationships between ideas and events are often stripped away.

Authors and editors do not willingly chop and flatten sentences, nor do they thoughtlessly mangle storylines.



The source of the writing problem is not in the publishing house, but in the public agency. Legislators, educational policymakers, and administrative regulators have unintentionally drained the life out of children's textbooks in several ways.

First, they have rejected textbooks that fail to achieve a mandated numerical score on a readability formula, a number they believe will guarantee a proper match between the text's difficulty and the reading ability of children at a given grade level.

Second, they have favored books that present a particular list of vocabulary words or teach a particular list of abstract skills (e.g. "finding the main idea") over books that might ignite children's imaginations.

Third, they have discouraged publishers from investing the time and effort it takes to produce carefully written material by failing to buy from publishers who do produce well-written textbooks.

READABILITY FORMULAS

Readability formulas were developed over 60 years ago to help educators choose textbooks written at the appropriate level of difficulty. The designers of readability formulas, then and now, operated on the assumption that long words, unfamiliar words, and long sentences are the primary causes of reading difficulty. Judging the difficulty of a text by such formulas is easy. Some formulas count syllables in words and words in sentences and calculate averages. Others rely on lists of words deemed to be familiar to children at various ages, adding points for words children are not believed to know. These mechanical labors yield a score that represents the grade level of a passage or a textbook.

When educators used readability formulas informally, along with common sense, to judge reading difficulty, these formulas had no harmful effects. But when policymakers and regulators began to make them part of official policies and procedures, the picture began to change. Publishers discovered they could lose a sale if an adoption committee subjected a randomly chosen passage to a formula analysis and found that the score was too high or too low. Defensively, textbook authors and editors began to write or adapt text so that it would survive a readability formula check. Short words ("it," for example) had to be substituted for long words ("elephant," for example).

The popular phrase, "dumbing down," aptly describes the effect of readability formulas on not only

To add confusion to boredom, short and vague words are often substituted for longer, precise ones. Instead of "esophagus," there will be "food tube." Instead of "protoplasm," there will be "stuff."

elementary books, but also junior high textbooks. Even the simple cadence of language can be cruelly monotonous. To add confusion to boredom, short and vague words are often substituted for longer, precise ones. Instead of "esophagus," there will be "food tube." Instead of "protoplasm," there will be "stuff."¹

Compound or complex sentences have been chopped in two, often at the expense of the reader's comprehension, in order to lower the score. For example:

ORIGINAL: "If given a chance before another fire comes, a tree will heal its own wounds by growing new bark over burned parts."

ADAPTED: "If given a chance before another fire comes, the tree will heal its own wounds. It will grow new bark over the burned parts."

In this instance, the way in which trees heal themselves—by growing new bark—is unstated in the edited version. "[A]n inexperienced reader, or one who does not know very much about trees, might make an incorrect guess and see healing wounds and growing new bark as separate processes, simply ordered in time," writes Andee Rubin, a textbook researcher.² This butchery is clearly attributable to an editorial effort to bring

down the readability score.

The limitations of readability formulas have been exposed by contemporary researchers. It is clear that readability formulas are blind to both meaning and style. Although their purpose was to provide a gross indication of grade level, the paradoxical effect of their misuse as a formula for writing has been to make text harder to understand. Moreover, the required use of words that children are presumed to know already puts too tight a rein on the growth of their vocabulary.

Despite the mounting evidence that writing to meet a formula is educationally unsound, specified formulas and scores are required in some states by law and in others by regulation. Even in states and localities with no formal requirements, judging the difficulty of a textbook by formula is still common practice.

Research has shown that at least five experienced teachers (five is the number needed to offset out biases and poor judgment), if given a modicum of training, are quite good at judging the reading difficulty of a text. Yet teachers have also become increasingly reluctant to exercise "subjective" judgment, preferring the safety of an "objective" measure. With so many dimensions of a book to consider in an adoption process, this one—Is the book easy enough?—can be nailed down without much effort. A committee member can check the "readabilities" with one eye on the TV or with the help of a computer program.

SKILLS MONGERING

Readability formulas are only one of the causes of bad writing. Another is the current accountability movement. Increasingly, educational accountability has become synonymous with student achievement on standardized tests of rudimentary skills. Few states or localities test subject knowledge, but virtually all

AS NORTH CAROLINA GOES, SO GOES THE NATION

W. Dallas Herring, a history buff and coffin maker from Rose Hill, North Carolina, was chairman of the North Carolina State Board of Education in 1966. In May of that year, when publishers were trying to get their U.S. history textbooks on North Carolina's state adoption list, Herring complained that North Carolina's role in U.S. history had been neglected.

According to Bill Graves, a reporter for the *Raleigh News And Observer*, Herring was particularly disturbed by textbooks that failed to mention the Revolutionary War battle of Moore's Creek Bridge, which lies about 45 miles south of Rose Hill. . . . Patriots in the 1775 battle stopped Loyalist troops from making a rendezvous on the coast

with British forces commanded by Gen. Charles Cornwallis.

In Herring's view, as reported by Graves in the *Raleigh News And Observer*, the battle was important "because it delayed Cornwallis' entrance into North Carolina and kept him from gaining an early foothold in the South."

Herring's fellow board members agreed with him and decided to put off adopting history books for a year. Harcourt, Brace and Jovanovich, whose *Rise of the American Nation* was the only book used in North Carolina at the time, produced a 1967 revision of the book with a full account of the Battle of Moore's Creek Bridge. Like many state schoolboard members, Herring was unaware of his power.

Graves reports that Herring "was surprised to see how responsive they were to my initial complaints." Said Herring, "It is a sensitive matter. After all, we are dealing with millions of dollars. And they were sweating blood."

Today, students across the nation still read a 250-word account of the battle in *Rise of the American Nation*, a book still widely used in North Carolina and elsewhere in the nation. "In fact," writes Graves, "they read more about the Battle of Moore's Creek Bridge than about the Boston Tea Party or the First Continental Congress. And the battle now appears in other U.S. history books as well." □

schools test the skills said to be critical to a mastery of content. Not surprisingly, textbook publishers have also begun to emphasize skills more than subject content.

Even when testing is limited to skills, the most available form of mass testing—the multiple-choice test—limits even skills to those that accommodate a short-answer form. Thus, even though writing is a “skill,” it will rarely be tested by asking students to write a composition, but by asking them to identify correct usage from among several options, to fill in blanks in sentences, to correct punctuation and capitalization, or other recognition tasks.

Similarly, reading is a “skill,” but the multiple-choice format shrinks most test passages to a paragraph. In the early grades, children are expected to learn letter sounds in certain sequences, and are tested accordingly on their ability to recognize similarities and differences in letter sounds in the test passages. Increasingly, textbooks are designed to help children pass such tests. “Too often,” says Andee Rubin, “the books for children in grades one through three are full of stories “whose claim to coherence is that they use the same vowel in

almost every word.”

Instead of designing a book from the standpoint of its subject or its capacity to capture the children’s imagination, editors are increasingly organizing elementary reading series around the content and timing of standardized tests. If commas are taught in September, but not tested until April, the book will administer a little comma dose every few lessons from October through May so children will be able to answer questions about commas on the test.

Frequently, says Diane Ravitch of Columbia University, the stories are being written by people who have never been heard of outside of textbook publishing houses. Under present policy, though, it could hardly be otherwise. Gifted writers of children’s trade books are reluctant to write for textbook companies, or even to surrender their already-published work to the technocratic editors concerned with readability formulas, vocabulary controls, and standardized multiple-choice tests.

The emphasis on testable skills is not limited to elementary textbooks. Books written for junior and senior

RECOMMENDATIONS

First and foremost, every adoption state governor, legislator, state board member, chief state school officer, and state agency official must recognize the paradox at the center of the textbook dilemma:

The harder they try to regulate the content of textbooks, the less useful textbooks become for their own students and students elsewhere.

POLICYMAKERS IN ADOPTION STATES SHOULD:

■ **Detach Their Curriculum and Testing Programs From Their Textbook Adoption Policies. Put Another Way, Adoption State Leaders Should Cease the Practice of Issuing Detailed, Skill-oriented Bid Specifications to Publishers.**

■ **Immediately Abandon the Use of Readability Formulas As an Adoption Criterion.**

■ **Stop Demanding a Recent Publication Date.**

■ **Appoint People to Adoption Committees and Selection Panels on the Basis of Their Knowledge and Talents, and Not**

Only on the Basis of Their Geographic Location, Their Institutional Role, or Their Organizational Affiliation.

■ **Provide Substantive Training for Adoption Committee Members.**

■ **Pay Teachers Who Serve on Adoption Committees, or Else Find Other Ways To Recognize and Reward Those Who Perform This Important Task.**

■ **Throw Away Long, Generic Checklists.**

Administrators might prefer the convenience of a checklist that can be applied to both horticulture and grammar textbooks, but the items that apply to both will almost certainly be limited to trivial matters and fail to capture important content-specific issues.

■ **Write It Themselves.**

If an adoption state is determined to have a textbook that covers every jot and tittle on its list of behavior objectives and test specifications, or feels strongly about respecting some powerful constituency with a strong ideological position, then it should let a sole-source contract to a reputable publishing house or a well-known

author who will treat that state’s curriculum coherently and in depth. There is ample precedent for this practice: Most states commission their own state histories.

■ **Devise Meaningful Sanctions Against State or Local School Personnel Who Seek or Accept Free Editions or Other Extras.**

■ **Establish and Implement a New, Subject-Specific Policy Framework for Curriculum Review and Textbook Selection With Two Central Objectives: Selecting Books on the Basis of Qualities That Are Known To Benefit Students, and Rewarding Publishers Who Produce Such Books.**

■ **Allocate State Funds for Textbook Purchases at the Local Level, and Earmark Some of the Funds for the Development of an Enhanced Selection Process.**

POLICYMAKERS IN NON-ADOPTION STATES SHOULD:

■ **Encourage by Their Purchasing Practices the More Venturesome and Less Expensive Offerings of Small Publishers, Use Trade Books**

high school students are laced with exercises claiming to develop "critical thinking skills" even though the text itself may do very little to stimulate thinking. Nobody would suggest that children should not learn how to think, but the dynamics of the textbook market encourage publishers to feature an almost content-free approach to "thinking," rather than allowing the content to drive children to think about what they are reading.

As these skills become imbedded in the "scope and sequence" charts, or "curriculum frameworks" of major states and cities, publishers are pressed to plan skills to be "taught" on each page, even before the text has been written. Publishers must also provide an index to the required skills, by page number, so that curriculum directors who don't have time actually to read the books can satisfy themselves that the skills are being "taught." A long string of page references behind "main idea," for example, impresses some curriculum directors who don't have time to discover that the main idea might not be worth finding. Clearly, many books are planned to satisfy the superficial selection process rather than to satisfy the curiosity of students.

THE "MENTIONING" PROBLEM

Another pervasive textbook sin is "mentioning," a term coined by researcher Dolores Durkin at the University of Illinois. The term refers to textbook prose that flits from fact to fact, statement to statement, and topic to topic, without giving the reader the context that would make sense of the factual information. Books accused of "mentioning" are generally long on facts and terms but short on ideas and explanations. Without the necessary context, readers often fail to see the significance of the connections between statements. Metaphors and similes, which would help readers grasp a complicated concept, are remarkably infrequent, even when we allow for how difficult it is to find metaphors that immature readers can understand. Examples and counterexamples that would give a concept some roundness are rare.

In an effort to satisfy the content requirements of so many adoption authorities, the text must be compressed into incomprehensibility. In science books, the

(Continued on page 39)

and Original Source Materials, and Otherwise Exercise the Freedom They Have.

School officials in non-adoption states can take many of the steps recommended to adoption states. In addition they should take greater advantage of the freedom they could enjoy. Unlike local school officials in adoption states, they are free to buy textbooks from small but worthy publishing houses; to substitute interesting trade books or supplementary materials for boring textbooks; to give—rather than loan—cheap paperbooks to students and buy new ones the following year. But for the most part, they don't do any of these things. Instead, they tend to limit themselves to the overstuffed, homogenized, expensive textbooks prepared for the big markets.

NATIONAL ACADEMIC ORGANIZATIONS OF TEACHERS AND/OR PROFESSORS IN THE VARIOUS DISCIPLINES SHOULD:

■ Define a Coherent Curriculum, or Several Alternative Coherent Curricula, That Can Be Taught by the Teachers We Have in the Time Allotted.

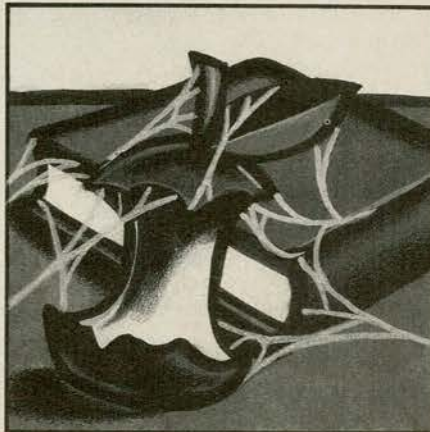
■ Those Responsible for Ethical Standards in Various Fields Should Specifically Discourage Professors and Teachers Who Agree To Be Listed As Authors (and Accept Royalties or Fees) for Textbooks They Did Not Actually Write.

■ Develop Model Contracts for Professors or Teachers Who Actually Do Write Textbooks. These Contracts Should Retain for the Author the Right To Control Subsequent Editions.

NATIONAL TEACHER UNIONS SHOULD:

■ Through Their Local Affiliates, Organize Content-specific Textbook Study Groups.

■ Promote Alternative Methods



for Judging the Reading Level of a Textbook.

■ As the Textbook Reform Movement Unfolds, Press for More Teacher Control Over Textbook Selection.

FOUNDATIONS SHOULD:

■ Support Independent, Critical Reviews of Textbooks in General Circulation Magazines and Newspapers.

PUBLISHERS SHOULD:

■ Employ True Subject-matter Scholars—as Opposed to Public School Teachers or Education Professors—to Review Textbook Content for Factual Accuracy, Conceptual Integrity, and Value.

■ Hire Authors Who Have Demonstrated Their Ability To Write Clear and Engaging Text for Mass Audiences.

■ Write or Select Good Material First, and Then Figure Out How To Use That Material To Foster Vocabulary Growth, Test Achievement, and Critical Thinking.

■ Become More Realistic About the Time It Takes To Produce a Good Book.

WHAT SHOULD YOUNG CHILDREN BE DOING?

Our kindergarten practices should reflect the best of what we know about how young children learn and develop.

BY LILIAN G. KATZ

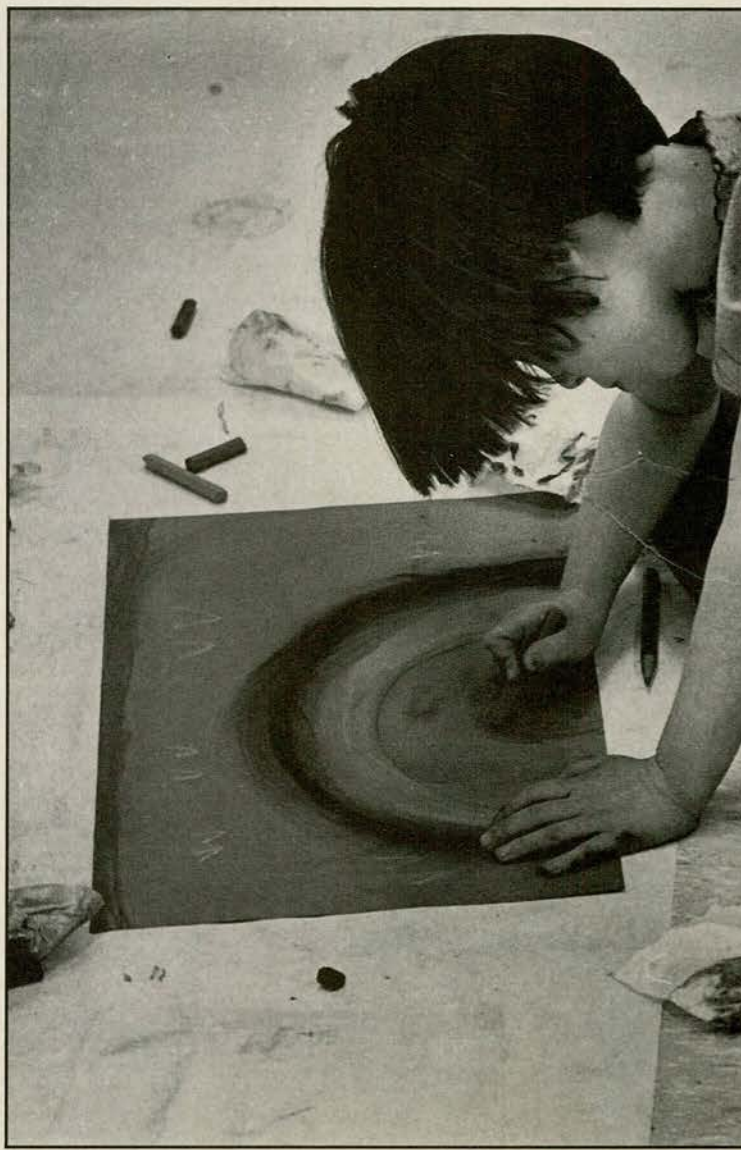
The founders of kindergarten envisioned it as a special place for the nurturance and guidance of the young child's spontaneous impulse to learn. Since the first American kindergarten opened its doors more than a century ago, disagreements about goals and methods have been common. Once again, in the 1980s, the role and functions of kindergarten have become a subject of sharp conflict and controversy in many communities. To make one's way through the tangle of issues currently under debate across the country, it is necessary to return to the basic question that must underlie all our thinking and all our policy decisions:

What should young children be doing, wherever they are: Whether at home with siblings or grandma, with a sitter, at a licensed family day care home, a Head Start class, a franchised day care center, Yale Child Development Center, or a rural or urban public school?

There is an abundance of research on intellectual and social development and learning, which is rich with implications for the kind of teaching and curriculum that should be provided for young children. Unfortunately, our practices are way behind what we know.

With respect to development, there are two meanings to the concept. One is what we call the normative dimension. It tells us something about what children can do at a given age or stage; what is typical, what is most frequently observed in children who are two and three and five and seven; at what age they know how many words, and at what age they can be expected to

Lilian G. Katz is Professor of Early Childhood Education at the University of Illinois (Champaign) where she also is Director of the ERIC Clearinghouse on Elementary and Early Childhood Education. This article is based on a presentation made by Dr. Katz as part of a program on early childhood education sponsored by the Johnson Foundation and The Prairie School in Racine, Wisconsin, and is adapted from a special insert to The Wingspread Journal, Volume 9, No. 2.





PHOTOGRAPH BY DENNE BRUNA

take their first step. When people say something is developmentally appropriate, they are usually referring to this normative aspect.

There is, in addition, another and very important meaning of the term development, and that is the dynamic dimension. It has to do with the sequence of learnings, the transformations that occur in capabilities from a given time period to another time period. The normative dimension tells us what most children can and cannot do at a given age. The dynamic dimension tells us what children should do at a certain time in light of the long-term consequences of early experience for later functioning.

Just because children can do something when they are young does not mean they should do it. It is well established within cultures and across cultures that young children can engage in a very wide variety of behavior. If you look at the developing world, you will see a large proportion of children at the age of six and seven engaged in child rearing, raising their siblings who are two and three, sometimes even younger. It can be done. That doesn't mean it should be done. Many of us have seen very young children selling newspapers at traffic lights and wondered how they manage to be so streetwise when they are so little. They can do it. You can see in day care centers and kindergartens young children working on worksheets or reading from flash cards. You can make children engage in rote counting of large numbers and do exercises reading the calendar. But that doesn't mean you should do it. You can make children work for gold stars and tokens and all sorts of rewards, but that doesn't mean you should. What's interesting is that almost anything you get young children to do they appear to be so willing to do. They don't appear to be suffering, and some of them even look as though they love it. The developmental question is not so much how children learn, because children always learn. They learn to lie, to steal, to mistrust, and so on. The develop-

mental question is what is it that young children should do that best serves their development in the long term.

FOUR CATEGORIES OF LEARNING

plain or tell us something. Knowledge is strengthened through studying and through repetition. Skills can be learned from instruction, from having directions, from having a coach. They can get better with practice and drill. Dispositions, however, are different. They are not learned through instruction or drill or lectures or workbooks. You can't very well give a child a workbook on curiosity and have it learned. There are many dispositions that we want children to acquire and strengthen—to be curious, creative, cooperative, friendly, helpful, hard working. These dispositions, however, cannot be learned from instruction. Dispositions are probably learned primarily from being around people who have them and who exhibit them. It is striking that there are dispositions like being curious that adults rarely exhibit in front of children, even in as simple a way as a teacher saying, quite sincerely, "I've been wondering if this is the best way to do such and such"—just to exhibit an inquisitive approach.

There are four categories of learning that I find helpful to look at. There may be ninety-nine, but let us focus on four.

Knowledge. This is a category of learning we're all concerned about. When it comes to young children, knowledge would include things like facts, concepts, stories. What always amazes me is how arbitrary the knowledge is that people select to teach children.

Skills. Skills are relatively small units of action that can be fairly easily observed and that occur in a relatively short period of time. There is a long list of skills. Somebody went to the trouble once to generate a list of skills for children and ran out of steam at 2,600. Skills would include motor skills and social skills, handwriting skills, counting skills. The list is almost endless.

Dispositions. Dispositions are a very different type of learning from skills and knowledge. They can be thought of as habits of mind, tendencies to respond to situations in certain ways. Curiosity is a disposition. It's not a skill, and it's not a piece of knowledge. It's a tendency to respond to your experience in a certain way. Friendliness is a disposition. Unfriendliness is a disposition. Creativity is perhaps a set of dispositions. Being bossy or a bully are dispositions. Not all dispositions are desirable.

Think about the difference between having reading skills and having the disposition to be a reader, or having writing skills in contrast to having the disposition to be a writer.

Feelings. The fourth category is feelings. Feelings are subjective emotional states, and a lot of them are probably innate, like feelings of fear. But some feelings are learned, and there are many feelings that concern educators, like feelings of confidence, feeling secure, feeling competent, and feeling as though you belong. These are important to us as teachers of young children and as parents as well.

Knowledge can be acquired by having somebody ex-



PHOTOGRAPH BY KAHNWEILER/JOHNSON

Dispositions are not learned through instruction or drill or lectures or workbooks. You can't very well give a child a workbook on curiosity and have it learned.

plain or tell us something. Knowledge is strengthened through studying and through repetition. Skills can be learned from instruction, from having directions, from having a coach. They can get better with practice and drill. Dispositions, however, are different. They are not learned through instruction or drill or lectures or workbooks. You can't very well give a child a workbook on curiosity and have it learned. There are many dispositions that we want children to acquire and strengthen—to be curious, creative, cooperative, friendly, helpful, hard working. These dispositions, however, cannot be learned from instruction. Dispositions are probably learned primarily from being around people who have them and who exhibit them. It is striking that there are dispositions like being curious that adults rarely exhibit in front of children, even in as simple a way as a teacher saying, quite sincerely, "I've been wondering if this is the best way to do such and such"—just to exhibit an inquisitive approach.

Another requirement for learning particular dispositions is to have the opportunity to behave that way, so that the behavior can be acknowledged and responded to. By acknowledging and appreciating a disposition, you strengthen it.

Children's dispositions can also be affected by the way we set tasks for them. For instance, if you say to children, "Today I want to see how good you are at something," you are setting a performance goal. The research indicates that when you set performance goals, the highly able children tend to say to themselves, "Good, I can show how smart I am." The children who are average or low in ability tend to say to themselves, "Am I going to show again that I don't have much ability?" They tend to become anxious, and their anxiety further deteriorates their performance.

On the other hand, you can set what are called learning goals in which you say to the children, "Today I want to see how much we can find out about such and such." When you set learning goals, children tend to see opportunities to achieve something, whatever their ability. Other research indicates that, with performance goals, children tend to become ego oriented and worry about how they will be judged; with learning goals, they tend to become task oriented, and the disposition to make an effort is strengthened.

Children can't all be equal in ability. Some children are more mathematically able than others. Some are more musical and some are more verbal. But they can come close to each other on effort. We ought, therefore, to say to children much more often than we do, "Let's see how much we can find out about something," rather than, "I want to see how good you are or how well you can do."

RISKS OF ACADEMIC PRESSURES ON YOUNG CHILDREN

In taking up this topic, it is useful to first distinguish between academic and intellectual rigor. *Academic* rigor refers to strong emphasis on completion of school-like tasks, exercises, grade-level achievement, grades and test scores, following instructions and meeting requirements, conforming to procedures and conduct

necessary to succeed in the academy and to fulfill its institutional requirements. *Academic* also suggests being out of touch and abstract. In contrast, *intellectual* rigor refers to characteristics of the life of the mind and its earnest quest for understanding, insight, knowledge, truth, solving intellectual puzzles, and the like.

Observation of young children in many preschool settings indicates that they can certainly be given instruction in phonics, counting, and other academic tasks. But the fact that children *can* do something is not sufficient justification for requiring it of them. While there is no compelling evidence to suggest that early introduction to academic work guarantees success in school in the long term, there are reasons to believe that it could be counterproductive.

Certainly young children can be successfully instructed in beginning reading skills; however, the risk of such early achievement is that in the process of instruction, given the amount of drill and practice required for success at an early age, children's dispositions to be readers will be undermined.

As we look at the results of longitudinal studies, the early pressure on young children to perform academic tasks introduced through direct instruction (for example, practice in phonics, workbook exercises) appears quite harmless, or even beneficial, in the short term. But we are obliged to take into account the potential long-term consequences of early experiences, no matter how benign they appear to be at the time they occur. Results from longitudinal studies suggest that curricula and teaching methods should be approached so as to optimize the acquisition of knowledge, skills, and desirable dispositions and feelings; and that these are mutually inclusive goals, each of equal weight. It is clearly not very useful to have skills if, in the process of acquiring them, the disposition to use them is lost. On the other hand, the disposition without the requisite skills is also not a desirable educational outcome.

What is sad to see in kindergartens is children so willing to do so many things that are so irrelevant to them at that age and so frivolous, and by second grade find many of them turned off. It isn't either-or. You don't acquire dispositions or skills. The challenge for educators—at every level—is to help the learner with *both* the acquisition of skills and the strengthening of desirable dispositions.

Another risk for preschool programs that emphasize academic or basic skills is that they tend to use a single teaching method and curriculum. Academically focused curricula typically adopt a single pedagogical method dominated by workbooks, drill, and practice. It is reasonable to assume that when a single teaching method is used for a diverse group of children, a significant proportion of them is likely to fail. It also seems to be a reasonable hypothesis that the younger the children are, the greater the variety of teaching methods should be. The younger the group, the less likely the children are to have been socialized into a particular and standard way of responding to their environments, and the more likely it is that the children's background experiences related to their readiness to learn are unique and idiosyncratic rather than common and shared.

For reasons of stability and practicality, there are, of

course, limits upon how varied the teaching methods can be. Approaches that are dominated by workbooks, drill, and practice often claim to individualize instruction. *What is typically individualized, however, is the day on which a given child completes a task, rather than the task itself!* Often "time on task" for children in such programs could be called "time on deadly task." After a year or two of such schooling, the effect on the disposition to learn is likely to be deadening!

Another risk that may attend introducing young children to academic work prematurely is that those children who cannot relate to the content or tasks required are likely to feel incompetent. When the content or tasks of a lesson for college students are difficult to grasp or perform, the student is very likely to fault the instructor, as many of us well know! However, in the case of young children—perhaps older ones as well—repeated experiences of being unable to relate to school work are likely to lead to the self-attribution of stupidity, which can be called "learned stupidity." Such children are then very likely to bring their behavior into line with this attribution.

I expect we can all remember at least one occasion in school when we thought we were the only person who didn't understand what the class was doing. Recall how that felt and then imagine feeling that way if you are four or five or six. Once probably won't hurt you, but if it happens over and over, it's like water dripping on a stone—it makes an indelible imprint.

THE DEVELOPMENT OF INTEREST

One of the important dispositions of concern to educators of young children is *interest*, or the capacity to "lose oneself" in an activity or concern outside of oneself. Interest refers to the capability of becoming deeply enough absorbed in something to pursue it over time, with sufficient commitment to accept the routine as well as novel aspects of work. This disposition appears to be present in the normal human at birth and is affected by a variety of social-psychological processes throughout childhood.

There are, however, many obstacles to acquiring the interest disposition in the television age where things occur in very short bursts of time.

Recent research has illuminated the effects of different kinds of feedback on learners' interest. Research on the so-called "overjustification effect" suggests that when children are rewarded for tasks in which they had initially shown interest, regardless of whether it's a verbal or concrete award, the reward is followed by loss of interest in the task. To the child the reward can signal that the task has no intrinsic value, and the child loses interest. Since this effect applies especially to those activities children originally find interesting, it suggests that teachers should exercise special care not to offer rewards for those activities young children spontaneously enjoy, find attractive, or are easily encouraged to engage in. This doesn't mean one shouldn't use rewards. What it means is that you have to use them extremely carefully.

A parallel line of research on related processes suggests that when the positive feedback given to children

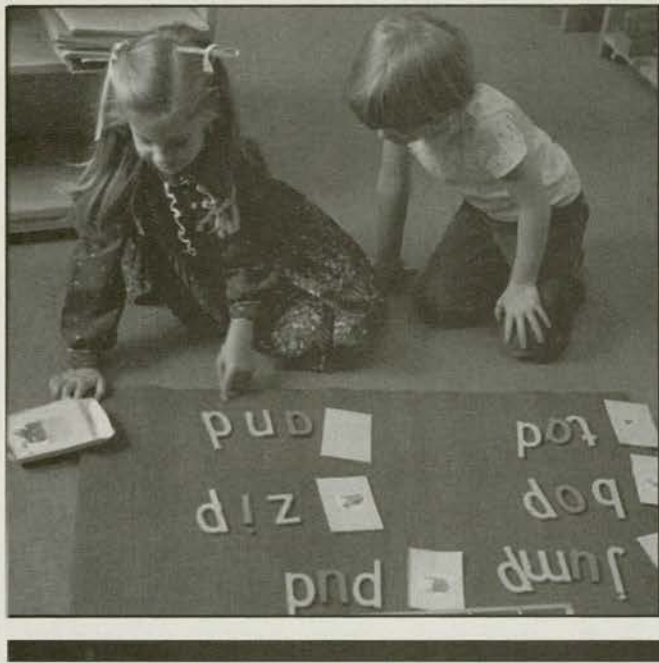
is general in nature, it may serve to increase productivity but not interest. General positive feedback includes vague comments on the part of the teacher like "very good," "well done," and the decorative smiling face or gold star. If, on the other hand, the positive feedback is specific rather than general, particularly if it includes information about the competence of the performance, it serves to strengthen interest. Academically oriented programs typically emphasize general positive feedback, ostensibly to give children feelings of success and to spur productivity. This strategy appears to work very well to induce young children to keep working at disembedded, decontextualized, and often very trivial tasks. However, the research on the effects of rewards strongly suggests that children may suffer academic "burnout" after two or three years of experience with general positive extrinsic rewards. If you just give general feedback, children tend to keep working and they increase their productivity, but they lose interest, which means when you stop saying, "it's great," they stop pursuing the task.

On the other hand, if you give positive feedback that is *informative*, that has information in it, they don't increase their productivity, but they do increase their interest, which means that they'll go on working when you're not there, and the disposition to go on learning is strengthened. The problem is that we have been taught in the last twenty years to overdo positive feedback. You can't keep up being informative every twenty seconds. What teachers need to do is keep the total amount of feedback low, but when they do give feedback, make it informative. Too many rewards can be distracting and intrusive and get in the way of children's spontaneous interest. Children need to engage in activities that call for sustained effort, extended work over time. This is a great problem in our elementary and early childhood curricula where the children's work is so fragmented. Too often teachers are saying, "Finished or not, it's time to go on to gym, music, computer, art, or whatever." The daily program for young children should allow for flexibility rather than fragmentation in the allocation of time to various activities.

Curricula and teaching methods that attempt to provide children with constant amusement, fun, and excitement also risk undermining the development of children's disposition for interest.

Another problem I notice is that I think we have confused emphasizing children's self-esteem with making them narcissists. That is, we do too much that turns children in on themselves. I've seen too many programs where children are wearing little signs that say, "I'm important." In one first grade class, a teacher had every child make a little book, "All about Me," with pages on what I like to eat, what I like to watch on television, what I want for a present, where I want to go on vacation. Every one of those pages was a consumer activity. There was no page that said, "What I want to learn more about." Our job as teachers is to engage children in deepening their understanding of relevant things outside of themselves. There are many ways to do this very early in programs for young children.

Thus, the teacher's role in strengthening children's dispositions to be interested in relevant and worthwhile phenomena is a complex and highly critical one. What is



PHOTOGRAPH BY ANN LICE

needed is for children to be encouraged to engage in projects that call for extension, elaboration, and continuation of work and play. Those are the kinds of experiences that cultivate the disposition of interest. The long-term consequences are of enormous proportions.

THE DEVELOPMENT OF COMMUNICATIVE COMPETENCE

Virtually all who are concerned with children in the early years recognize that early childhood is a critical period in the development of communicative competence, namely competence in self-expression and in understanding others.

Contemporary insights into the development of communicative competence in young children indicate that all three basic functions of language (communication, expression, and reasoning) are strengthened when children are engaged in conversation, rather than when they are simply passively exposed to language. Just being exposed to language is not enough. In fact, our children are exposed to a great deal of language, some of which we wish they weren't. But it's conversation that strengthens all the communicative competencies.

Conversations are a very special type of interaction in which the content of each participant's contribution is contingent upon the contributions of others, in a sequential string of interactions. It may very well be that the *contingency* of adults' responses to children in and of itself has a powerful effect on the development of their intellects. Conversation is more likely to be prolonged when adults make comments to children than when they ask them questions. Too much classroom activity is one-way communication, teacher to class.

What we are learning is that the probability of conversation increases when children are in small groups of three or four, with or without an adult present. Most teachers of young children recognize the difficulty of encouraging conversation during a whole group session; they expend much effort reminding children that another child is still speaking or that their turn has not yet come! We also know that the probability of conversation increases when there's something of interest to talk about, something that has meaning to the lives and the experience of the participants. I watched a kindergarten teacher who was attempting to engage a class of five-year-old children in a discussion by asking each in turn, "What is your news today?" Each child struggled to find something headline worthy to report to his or her disinterested squirming classmates! Perhaps some of these children were learning "to listen," but many appeared to be learning "to tune out" their stammering classmates.

What helps in conversation is when teachers say, "Yes, nice point, that's true," or something of that sort, not heavy or phony, but genuinely reflecting interest in what the child is saying.

SOCIAL COMPETENCE

The evidence is now very compelling that young children who do not have minimal social competence
(Continued on page 44)

What is typically individualized, however, is the day on which a given child completes a task, rather than the task itself!

FLUNKING KINDERGARTEN: ESCALATING CURRICULUM LEAVES MANY BEHIND

BY LORRIE A. SHEPARD AND MARY LEE SMITH

NEXT YEAR Michael Lee will repeat kindergarten because he flunked the Georgia readiness test for first grade. Judging from his fidgety behavior and inability to cope with scissors and other fine-motor tasks, his teacher confirmed that Michael Lee could benefit from another year before going on to first. Across the country the practice of kindergarten retention for many children like Michael Lee is increasing dramatically. In some districts, 10%, 25%, 33%, or as many as 60% of kindergartners are judged to be unready for the academic rigors of first grade. Unready children are provided alternative programming: developmental kindergarten (followed by regular kindergarten), transition or pre-first grade, or repeating kindergarten.

An extra year before first grade is intended to protect unready children from entering too soon into a demanding academic environment where, it is thought, they will almost surely experience failure. Depending on the philosophical basis of kindergarten retention, which differs profoundly from one district to the next, the extra year is meant either to be a time for immature children to grow and develop learning readiness or a time to work on deficient prereading skills.

The advocates of kindergarten retention are undoubtedly well intentioned. They see retention as a way for the school to respond to the enormous differences in background experiences, developmental stages, and aptitudes of the young children who present themselves at the schoolhouse door. They view it as a policy that has the child's best interests at heart and as a means, as they would put it, to prevent failure before it occurs. The question is: Are they right? Is kindergarten retention a helpful remedy for the problems it is intended to address?

For the past four years we have conducted research on the issues surrounding kindergarten retention: What are current practices? What problems are encountered

by children who are youngest in their grade? How accurate are the tests used for screening? What are the effects of extra-year programs? What are the differences in school cultures that account for low incidence of retention in one school and high incidence in the next?

In this article we summarize three of our major findings: 1) Kindergarten retention does nothing to boost subsequent academic achievement. 2) Regardless of what it is called, the extra year creates a social stigma. 3) And most ironically, the practice of kindergarten retention actually fosters the problem it was intended to solve—it feeds the escalation of inappropriate academic demand in first grade.

We have been able to locate 14 controlled studies that document the effects of kindergarten retention: 6 studies that were included in Gredler's 1984 major review of the research on transition rooms and 8 newly identified empirical studies.¹ The dominant finding is one of no difference. Gredler concluded that at-risk children, promoted to first grade, achieved as well or better than children who spent an extra year in transition rooms. The additional studies we located confirmed Gredler's conclusion. Children who spend an extra year in transition rooms are no better off at the end of first grade than comparable children who were recommended to repeat but whose parents refused.

In the study we conducted in Colorado, extra-year children were matched with control children on sex (mostly boys), birth month (mostly near the entrance-age cutoff), and measured readiness at the start of kindergarten. Measured at the end of first grade, there was again no difference on standardized math scores or on teacher ratings of reading and math achievement, learner self-concept, social maturity, and attention span. The only significant result was on standardized reading scores, where the retained children were only one month ahead of promoted children. These no-difference findings are surprising considering that by this time the retained children were a year older and had had an additional year of schooling compared to the control children who began equally at risk. That is, each group of children was tested at the end of first grade. But by the time the retained children were tested, they had had two years of kindergarten and one year of first grade, as compared to one year of kindergarten and one year of first grade for the promoted children.

When parents are asked to agree to retention or transition placement they are often told that, given the

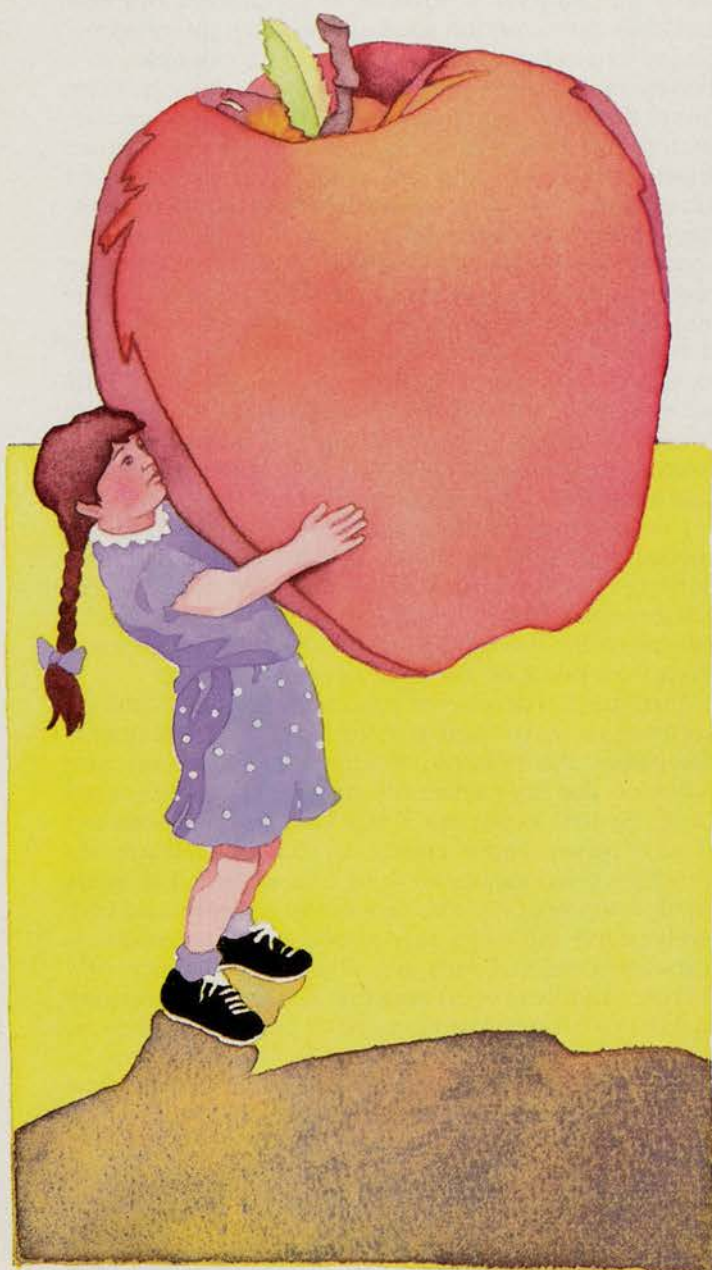
Lorrie Shepard is chair of Research and Evaluation Methodology in the School of Education at the University of Colorado, Boulder. Her research focuses on uses of tests in educational settings. Mary Lee Smith is professor in the College of Education, Arizona State University. She specializes in qualitative research methods and studies of diversions in pupils' careers.

extra year to grow, their children will move to the top of their class and become leaders. Research evidence from controlled studies does not support this claim.

How could there be such a discrepancy between research findings and the practical experience of many teachers who watch children blossom and grow during their transition year? For example, a study conducted by Dr. Judith Ford in Norman, Okla., is often cited by the Gesell Institute to support its advocacy of extra-year programs.² During their year in transition class, the 27 children in the Norman program gained an average of 55 percentile points on the Metropolitan Readiness Test. Thus children who were in the bottom half of their class at the end of one year of kindergarten were remarkably more ready after an extra year, now with readiness scores more like those of their more mature peers who had gone directly on to first grade.

Though many cite findings such as these as convincing, this study is fatally flawed. As is typical of studies cited by transition advocates, the Norman study had no control groups, which would have been critical in determining what those children would have been like if they had been promoted rather than retained or placed in transition. Nor were children in the Norman study followed up in first grade. Studies with control groups consistently show that gains such as these in readiness *do not persist into the next grade*. Eventually children end up at approximately the same percentile rank compared to their new grade peers as they would have been had they stayed with their age peers. Young or at-risk students who are promoted perform equally well in first grade.

Kindergarten teachers, however, are generally unaware of these end results. They know only that the retained children are doing better than they did in their first year of kindergarten. In the short run, teachers see progress: longer attention spans, better compliance with classroom rules, and success with paper-and-pencil tasks that were a struggle the year before. Furthermore, many of the transition children are above-average achievers in their first grade class (but, unseen by their teachers, so are an equal number of the matched control children). Some of the transition children are still acting out and doing poorly with worksheets (as are an equal number of control children). After retention has been tried and children are a year older than their classmates, disruptive behaviors that were once thought to be signs



of immaturity are now seen as relatively enduring personality traits.

For these few transitory academic benefits, retained children pay with a year of their lives. And, they understand that they could not go on with their classmates because of something that was *wrong with them*. Many educators believe there is no stigma attached to kindergarten retention, especially if it is "handled properly" by parents. Many especially deny that transition placement—which has a different name and does not involve recycling of curriculum—could be harmful. But children know that they are not making normal progress in the same way they know the meaning of placement in the bluebird reading group. One little girl understood the meaning of her pre-first placement so well that she thought she would also need to go to pre-second before second grade, and pre-third, and so on.

Our conclusion that kindergarten retention is traumatic and disruptive for children is based on interviews at the end of first grade with parents who had previously agreed to developmental or transition placements for their children. The majority of parents said that on balance the extra year had been the correct decision. Even if their children were doing poorly in first grade, they believed they were ahead of where they would have been without the extra year (and we did not tell them that the control group made equal progress). A majority of parents also reported significant negative emotional effects associated with the retention. The apparent contradiction was created by the substantial group in the middle who reported both positive and negative experiences.

Children know that they are not making normal progress in the same way they know the meaning of placement in the bluebird reading group.

The following quotations typify the ambivalent feelings of parents who gave a positive "vote" to the program but revealed an undercurrent of regret:

I knew he was struggling and he knew that he wasn't doing what the other kids were doing so I thought this was right. He's OK now. He does refer back once in a while. He says, "If I would have made it through kindergarten, I would be in second grade instead of first."

Well, the only [problem] was that he wasn't going to be going with the rest of his class into the next grade. But it was only because I told him that he was so special that his teacher wanted to keep him.

I think the biggest drawback is the attitude of other children and adults. Not so much from the teachers, but parents of other children remarking on how he looks so

much older, "he should be here," "he should be there," and other children picking up on the fact that he was going to remain in kindergarten, giving him a hard time about that.

I think it was more of a social thing. It was really hard to explain to her that her friends would be going on and she wouldn't be. That was a real hard part of it. I think it helped her more than it hindered her.

I personally think it's better that we've held her back and she has the possibility of being a little closer to the top than being a grade ahead and being at the very bottom. Some of the negative aspects of it are her own problems dealing with it and saying that she's been held back.

Kindergarten retention is similar in many respects to tracking and special education placements for mild learning problems. The logic of providing instruction tailored to individual learning needs is admirable, but research has not confirmed the efficacy of separate placements. Instead, research has documented negative side effects such as social stigma, lowered expectations, and watered-down instruction. From findings in these other literatures it is possible to speculate about why kindergarten retention does not produce the expected boost in academic achievement. For example, in a review of research on ability grouping, Robert Slavin found that homogeneously tracked classrooms are ineffective but that within-class groupings do improve learning.³ He reasoned that within-class grouping for each subject provides a closer fit between student learning and instruction than does a one-time assignment to separate classes on the basis of ability. Similarly, we might reason that kindergarten retention is a very gross and inaccurate way to individualize instruction because it requires a 12-month dislocation. Children who are judged to be unready by three months are treated the same as those who are 12 months behind; a child who seems immature in only one area of development is treated the same as a child who suffers developmental lags in all areas of development.

Kindergarten retention also resembles tracking and special education placement in that a disproportionate number of minority children are selected for extra-year placements.⁴ Thus, children who most rely on public education for the opportunity to learn are segregated from their peers on the basis of prior learnings.

Tests used to determine readiness are not sufficiently accurate to make their use for extra-year placements defensible. For example, Kaufman and Kaufman have provided the only reliability data on the widely used Gesell School Readiness Test.⁵ They found a standard error of measurement equivalent to *six months*, meaning that a child measured to be four and one-half years old developmentally and unready for school could very likely be five and fully ready. Although various readiness tests are correlated with later school performance, predictive validities for all available tests are low enough that 30 to 50 percent or more of children said to be unready will be falsely identified.⁶

Over the long term, kindergarten retention has a final negative consequence. Children who are over age for their grade have a much greater likelihood of dropping out of school. The Association of California Urban School Districts reported that children failed in their first two years of school have substantially reduced chances of completing high school.⁷ When background

The more that unready children are screened out of school or put in pre-K, the more that kindergarten becomes a place for six-year-olds.

factors and achievement are taken into account, children who have been retained or are otherwise over age for their grade are 20 to 30 percent more likely to drop out.⁸ These findings hold true in both rich and poor school districts.

THE CURRENT fad to flunk children in kindergarten is the product of inappropriate curriculum. Over the past 20 years there has been a persistent escalation of academic demand in kindergarten and first grade. What were formerly next-grade expectations are shoved downward into the lower grade. In a recent survey, 18 percent of principals reported that it is district policy to teach reading to all kindergartners; an additional 50 percent of schools teach reading to kindergartners who are "ready and able"; 85 percent of elementary principals say that academic achievement in kindergarten has medium or high priority in their schools.⁹

In a forthcoming article for the *Elementary School Journal*, we document the societal factors behind the escalation: universal availability of kindergartens, pervasiveness of preschool, and Sesame Street. If everyone has had kindergarten, then first grade teachers assume as prerequisites those letter sounds that previously were taught in first grade. If kindergartners already know their letters from Big Bird, then they must be taught something more, or so the argument goes. In addition, our interviews with teachers reveal more immediate sources of pressure: accountability gates in later grades and demands from middle-class parents that children move faster and faster along the track of pre-primers and graded workbooks. Schools with high rates of retention in kindergarten are characterized by an "accountability culture." Promotional gates at third grade or sixth grade are translated downward into fixed requirements for the end of first grade. If a first grade teacher is visited by the principal and reprimanded for any child who is below national norms on standardized tests, this teacher in turn communicates to the kindergarten teacher an unwillingness to accept children for first grade who are not ready to read.

Kindergarten teachers also describe the demands imposed by parents. Many middle-class parents visit school and convey that their only criterion for judging a teacher's effectiveness is her success in advancing their child's reading accomplishments. They ignore other evidence of enriching experience and cognitive development. "My child was reading when he came to school. You haven't taught him a thing." What counts for many parents is the number of first grade primers completed

in kindergarten because this is a clearly quantifiable measure of progress, like an SAT score for a five-year-old.

More academics borrowed from the next grade is not necessarily better learning. A dozen national organizations, such as the National Association for the Education of Young Children, the International Reading Association, and the National Association of Elementary School Principals, have issued position statements decrying the negative effects of narrow focus on literacy and numeracy in the earliest grades.¹⁰ Long hours of drill-and-practice on isolated skills are detrimental to all children, even those who are able to meet the demands, because tiny, boring proficiencies learned by rote are substituted for conceptual understanding and enthusiasm for learning. Highly formalized activities that occur too early deprive children of time to learn from play, substitute inappropriate symbolic learning for manipulative learning, detach reading from normal language development, stifle natural exploration, and increase stress.¹¹ More seriously, fixed, higher standards injure at-risk pupils, causing many more children to fail who would have, in due course, done quite well. The clearest victims of inappropriate curriculum are the children who are judged inadequate by its standards, children who can't stay in the lines and sit still long enough.

Many kindergarten teachers acknowledge that extra-year programs would not be necessary if children were being sent on to a more flexible, child-centered first grade. But faced with what they expect will be a punishing experience for the child (and holding generally rosy opinions about the effects of retention), keeping the child in the safety of kindergarten is clearly preferred. Educators do not express awareness, however, that the practice of retention might actually contribute to the escalation of curriculum. The more that unready children are screened out of school or put in pre-K, the more that kindergarten becomes a place for six-year-olds. Teachers naturally adjust what they teach to the level of the children in their class. If many of the children are older and reading, teachers do not continue to teach as if the room were filled with five-year-olds. Likewise, as more and more "unready" children are removed, first grade becomes a place for seven-year-olds, and instruction is paced accordingly. The subtle adjustment of curricular expectations to the capabilities of an older, faster-moving group can be demonstrated in the research literature on school entrance ages.¹² Each time a district or state raises the cutoff date for school entry, the hope is to eliminate the youngest children who seem unready for school. In a very short time, instruction is adjusted to the new age range and a new youngest group appears inadequate.

One alternative to escalation, retention, and more escalation can be found in the schools we observed that practiced virtually no kindergarten retention. Instead of highly stratified curricula, strict promotion standards, and an insistence that teachers adhere rigidly to the authorized curriculum rather than exercising their creativity, these schools had developed a culture where teachers and principal shared a commitment to adapting curriculum and instructional practices to a wide range of individual differences. They were able to manage heterogeneity without the need to

sort, label, track, and retain. Although these non-retention schools were also very academic and teachers had goals for skill development in kindergarten, a child who was not yet proficient would not be failed. Instead there were cooperative understandings between teachers. The kindergarten teacher would begin at the child's level and move him along to the extent possible, and the first grade teacher would pick up where the kindergarten teacher left off. These schools also had more flexible between-grade arrangements. Children moved more freely across grade boundaries, as exemplified by cross-age tutoring or a child visiting the next-higher grade three hours a week for reading instruction.

We are told that all of our conclusions are credible except the implication that current practices can be changed.

Our observations indicated that the non-retention schools were neither richer nor poorer than those schools with rigid grade-level expectations; nor did they serve less diverse populations. It should also be noted that the more flexible and individualized arrangements in the non-retention schools did not come at the expense of higher standards. The average standardized achievement test scores for third graders in these schools were no different from those in the high-retaining schools that had become preoccupied with the accountability tests.

When these research findings are presented to groups of educators across the country, we are told that all of our conclusions are credible except the implication that current practices can be changed. A kindergarten teacher stands up in the audience and gives yet another account of what will happen to children who cannot keep pace in first grade. In a workshop for first grade teachers the story is told of the principal who visits each May, test scores in hand, seeking an explanation as to why several of the children are not above national norms. In a state conference of elementary principals, the principals point to their superintendents, who post standardized test scores by school. As long as each group feels powerless to intervene and persists in practices that contribute to the problem, the problem will get worse. More and more children like Michael Lee in Georgia will be told, in one of their earliest encounters with schooling, that they are inadequate.

The answer is still to be found in the schools with appropriate curriculum and collegial understandings among teachers and principal that make retention unnecessary. Once the larger context of curriculum escalation is understood, then perhaps groups of early-grade teachers and their principal will have greater incentive to resist the myriad pressures and reject the factory-model, accountability culture that is rendering more and more children "unready." □

REFERENCES

1. Gredler, G.R.: Transition classes: A viable alternative for the at-risk child? *Psychology in the Schools* 21: 463-470, 1984; Dolan, L.: A follow-up evaluation of a transition class program for children with school and learning readiness problems. *The Exceptional Child* 29: 101-110, 1982; Jones, R.: *The effect of a transition program on low achieving kindergarten students when entering first grade*. Unpublished doctoral dissertation, Northern Arizona University, 1985; Kilby, G.A.: *An ex post facto evaluation of the junior first grade program in Sioux Falls, South Dakota*. Unpublished doctoral dissertation, University of South Dakota, 1982; Shepard, L.A.: Evaluation of transition room. Prepared for Kirkwood School District, January 1984; May, D.C., and Welch, E.L.: The effects of developmental placement and early retention on children's later scores on standardized tests. *Psychology in the Schools* 21: 381-385, 1984; Mossburg, J.W.: *The effects of transition room placement on selected achievement variables and readiness for middle school*. Unpublished doctoral dissertation, Ball State University, 1987; Stapleford, D.C.: *The effects of a second year in kindergarten on later school achievement and self-concept*. Unpublished doctoral dissertation, Michigan State University, 1982; Turley, C.C.: *A study of elementary school children for whom a second year of kindergarten was recommended*. Unpublished doctoral dissertation, University of San Francisco, 1979.
2. Ford, J.: *The Transitional Classes: A Report to the Elementary Principals*. Norman, OK, Norman Public Schools, October 22, 1985 (mimeographed).
3. Slavin, R.E.: Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research* 57: 293-336, 1987.
4. California School Readiness Task Force: *Here They Come: Ready or Not*. Sacramento, California State Department of Education, February 1988.
5. Kaufman, A.S., and Kaufman, N.L.: Tests built from Piaget's and Gesell's tasks as predictors of first-grade achievement. *Child Development* 43: 521-535, 1972.
6. Shepard, L.A., and Smith, M.L.: Synthesis of research on school readiness and kindergarten retention. *Educational Leadership* 44 (3): 78-86, 1986.
7. Shepard, L.A.: *Dropouts from California's Urban School Districts: Who Are They? How Do We Count Them? How Can We Hold Them (Or at Least Educate Them)?* Association of California Urban School Districts, October 1985.
8. Grissom, J.B.: *Structural Equation Modeling of Retention and Overage Effects on Dropping Out* (in press).
9. Educational Research Service: Kindergarten programs and practices in public schools. *Principal*, May 22-23, 1986.
10. National Association for the Education of Young Children: NAEYC position statement on developmentally appropriate practice in the primary grades, serving 5- through 8-year-olds. *Young Children* 43 (2): 64-84, 1988; International Reading Association: Literacy development and pre-first grade. *Childhood Education* 63: 110-111, 1986.
11. Elkind, D.: *Miseducation: Preschoolers at risk*. New York, Alfred A. Knopf, 1987; International Reading Association: Literacy development and pre-first grade: A joint statement of concerns about present practices in pre-first grade reading instruction and recommendations for improvement. *Childhood Education* 63: 110-111, 1986; Kamii, C.: Leading primary education toward excellence: Beyond worksheets and drill. *Young Children* 40 (6): 3-9, 1985; National Association for the Education of Young Children: NAEYC position statement on developmentally appropriate practice in early childhood programs serving children from birth through age 8. *Young Children* 41 (6): 4-29, 1986; Winn, M.: *Children without childhood*. New York, Random House, 1983.
12. Shepard, L.A., and Smith, M.L.: Escalating academic demand in kindergarten: Counterproductive policies. *Elementary School Journal*, Fall 1988 (in press).

TEXTBOOK FIASCO

(Continued from page 27)

density of new, italicized (but poorly explained) technical terms on each page is a good measure of the extent of mentioning. Entire books, like the biology example below, are often glossaries masquerading as textbooks.

NUCLEIC ACIDS New vocabulary: chromosome, nucleic acid, DNA, RNA, nucleotide.

In the nucleus of a cell are threadlike strands called *chromosomes*, (KRO-muh-somz). They are composed of proteins and *nucleic acids* (noo-KLAY'-ik). The proteins in nucleic acids make up two important chemicals, *DNA* and *RNA*. Nucleic acids are organic compounds that are made up of carbon, hydrogen, oxygen, nitrogen, and phosphorus.

DNA and RNA are not the only nucleic acids, but they do have special roles in the cell. RNA is involved in making proteins. DNA is involved in controlling the cell's activities. *Both are involved in passing on characteristics from parents to offspring.*

Each nucleic acid is made up of units called *nucleotides* (NOO'-klee-uh-tidz). In turn, each nucleotide is composed of three parts: a chemical group containing phosphorus, a group containing nitrogen, and a simple sugar.

If you find this incomprehensible, pity the poor ninth grader. In this tangle of passive voice sentences, cause-and effect relationships become lost. The author switches back and forth between parts and chemical compounds without warning. The signals—"are composed of" and "are made up of"—are inconsistently applied. The intelligent response to such "mentioning" and bad writing is "So what?" or "Who cares?"

The "mentioning" problem, like the bad writing problem, is directly attributable to public policies and procedures. Adoption states that generate excessively detailed textbook specifications seldom take into account the time it would take to teach all their required items, or the space available in a standard-sized textbook. Typical selection procedures seldom take into account the critical mass of information a student needs to understand an unfamiliar topic.

The problem of too many topics in too little space is especially severe in social studies, history, and science books. The Thirty Years' War will be "covered" in a paragraph; the Nixon presidency in two sentences. Nucleotides will be mentioned, and the glossary will contain a circular definition, but the student will not learn much about them. All of the small facts and terms that can be tested on a multiple-choice test will appear in the index, because that is where adoption committees usually check on curricular and test "congruence"—if they check at all.

In recent decades, the "mentioning" problem has become more acute. Special-interest groups pressure policymakers to include more material in the curriculum (and therefore the textbooks) about their favorite subjects. Policymakers find it difficult to resist these pressures because, for the most part, the additions sound reasonable. A state or local schoolboard can submit, without a troubled conscience, to demands from environmentalists, the health food lobby, advocates of the work ethic, and any organized minority group.

Even where good causes are not involved, there are adult pressures to teach more and more academic mate-

rial as the scope of knowledge within disciplines expands. School systems, test developers, and textbook publishers often ask university professors to serve on advisory committees, and in that setting, professors generally defer to one another, cheerfully adding each other's suggestions to the list of what should be taught.

With so much to stuff into the book, editors make sacrifices. Since publishers are held to account for a jumble of topics and facts, but not for coherence, coherence suffers. A thoughtful reader finds it tough to detect the pattern that has determined an author's choices.

Lacking any firm basis for choosing material, and required to include so much, textbook authors easily fall into the "mentioning" trap. A student may be told, for example, that Aristotle "studied the political organization of 150 city states and put down his conclusions in a book called *Politics*." He won't be told, however, what Aristotle's conclusions were.

At the moment, school officials prefer mentioning to coherence because they are obsessed with the idea that the textbook must cover as many of the facts and topics in the curriculum and tests as possible. With so little time to examine books, adoption committees check up on textbook/curriculum/test congruence by checking the labels, captions, index, and glossary. Knowing how superficially books are examined, publishers are best advised to sacrifice depth and comprehensibility and concentrate on coverage, however inadequate it may be.

Publishers also sacrifice material that may cause them to be criticized or to lose sales. Pressures from the politically organized, religious right have made it risky for publishers to discuss evolution. If evolution is discussed at all, it is often confined to a chapter at the end of the book. Students are conducted on a forced march through the phyla, and given no understanding of the overarching theory (evolution) that gives taxonomy life and meaning. Touchy subjects, like dinosaurs, the fossil record, genetics, natural selection, or even the scientific meanings of the words "theory" and "belief" are treated skimpily or vaguely in order to avoid fundamentalist ire.

Bad writing and the "mentioning" problem are intimately related. It is hard to write well about a vast span of history in one paragraph. A scientist might call a one-page explanation of photosynthesis "inaccurate" while a writer will call it "badly written." They are both right, but they have examined the text from different perspectives. Sense and style are intimately related, and so are space and accuracy, as every newspaper reporter know.

Some teachers defend today's outline-style textbook on the grounds that they can fill in whatever information the textbook omits. Such a defense suggests that the book is not even expected to be comprehensible on its own. Many teachers no longer see the book as material for students to read, but as a reference guide to the material that is supposed to be covered in class. They have, in effect, given up on the possibility that a textbook can be an independent source of learning. □

REFERENCES

¹Armbruster, Bonnie B., J. Osborn, and A. Davison: Readability formulas may be dangerous to your textbooks, *Educational Leadership*, April 1985, pp. 18-20.

²Rubin, Andee: What can readability formulas tell us about text? in R.C. Anderson and J. Osborn (eds.): *Foundations for a Literate America*. Lexington, MA, Lexington Books, 1984.

(Continued from page 22)

tendent of the Chipville School District, an affluent suburb of the capital city, Mylanta, is named to the task force, along with two experienced teachers from rural districts.

The two teachers, much to everybody's expressed regret, cannot attend the first meeting of the task force because there is no money in their local districts' budgets for release time. Nevertheless, the attending members put forth an eclectic array of philosophies about subject matter and pedagogy, and everyone has strong feelings about the particulars children should know.

Harvey and Helmut are pleased when Dr. Henry from Chipville says that his community's business leaders are concerned about the "bottom line." Dr. Hamilton tosses in some impressive suggestions that nobody understands.

Nobody at the first meeting has taught school in a number of years, and Helmut has never taught at all. His comments about testing, though, are respected because his specialty requires so many years of schooling. Harvey writes down comments, making no attempt to resolve conflicts or discrepancies, because he wants to respect everyone's input.

Before the second meeting of the task force, Harvey's superior, the Assistant Superintendent, reads the minutes of the first meeting and warns Harvey that a task force report without teacher input is likely to encounter criticism from members of the State Board. So Harvey finds state funds to pay for substitutes so the two teachers can attend the next meeting.

AT THE SECOND meeting, everybody listens to the list of goals and objectives that Harvey distilled from the previous discussion. At first, the teachers are afraid to say anything in the presence of so many experts. Eventually, though, Miss Hill makes a plea for a less-congested curriculum. "There's more content on this list than my kids can get through in *twelve* years of school, let alone six," says Hill.

Dr. Henry counters with a mini-sermon about rigor and high expectations. "We have to get past this idea that the kids can't learn very much," he says. "There are low-track kids in Chipville who are doing very well in the third grade with a college prep, fast-paced curriculum."

The other teacher, Miss Holdren, is seething with rage, but speaks up in a polite, high-pitched voice. "College prep doesn't mean very much to kids who just arrived in this country," she says. "My kids barely speak English, and they need a curriculum that respects the culture they have left and also teaches them practical things, like how to make change in American money at the supermarket." None of the men on the task force wants to oppose a practical idea, so nobody says anything. But Harvey doesn't write down her comments because he thinks Holdren is too sentimental for the schools of the 1980s, and secretly suspects her of closet bilingualism.

After the second meeting of the task force, Harvey and Helmut whip the input into shape for the State Board of Education's monthly meeting the following Tuesday. Helmut converts the views of the task force members—

which he thinks were rather loosely expressed—into precise behavioral objectives that correspond to the items on the statewide standardized test used in Nirvana. As he teases the content into test-sized cubicles, some of the larger ideas expressed by the task force members get lost.

Also, because Helmut doesn't want to dictate the details of the content to local school districts, he concentrates on defining the skills that every Nirvana student should know, regardless of the content used to teach it. It isn't very important, he thinks, whether kids study Greece or Mexico as long as they learn how to find the main idea, how to distinguish fact from opinion, and how to make correct inferences.

As he teases the content into test-sized cubicles, some of the larger ideas expressed by the task force members get lost.

Harvey and Helmut do take one aspect of curricular content seriously: a number of national commissions, along with Dr. Hamilton, have spoken out on the need for students to know something about the history of science, especially realistic and inspirational biographies of scientists. They assemble a list of 127 scientists—from Agassiz to Young—that elementary children should study. But since there are so many scientists, only two sentences can be devoted to each. The first sentence identifies the discovery that made the scientist famous, and the second sentence—the one in which Harvey takes special pride—is designed to give the scientist a "human touch." Thus students will be expected to memorize the fact that Svante Arrhenius's teachers gave him the lowest possible passing grades and that Wallace Carothers killed himself at 41 after his twin sister died.*

Responding to the teachers' concerns about books that are too hard for some students to read, Harvey tucks in a requirement that those adopted in Nirvana must be .2 below grade level on the Haraugh Readability Index, the formula used by the state's test contractor. He and Helmut feel good about the high degree of congruence they have achieved between the curriculum, the tests, and the textbook specifications.

HARVEY'S SUPERIOR, who wasn't present at either of the task force meetings, presents the draft curriculum/bid specifications to the Board. Half of the members have not read the document prior to the meeting because they think that curriculum and textbooks are really professional matters rather than policy concerns.

But the member from Novisa City worries aloud about whether the students in his district can benefit from books written in English since most of them speak

*Actual examples from California's *Science Model Curriculum Guide, Kindergarten Through Grade Eight*, 1987.

only Spanish at home. On a four-to-three vote, the Board approves his motion to require publishers to produce a Spanish edition if they intend to submit textbooks for adoption in Nirvana.

The member from Mylanta, who is a local history buff, has always been irritated by the fact that textbook publishers don't even mention an important Civil War battle, which took place just outside Mylanta and was critically important, in his view, to the outcome of the Civil War. On his motion, the Board adds this battle to the list of topics to be covered in the textbook series. The member from Chipville, a prominent businessman, says his employees can't solve even simple problems and asks that "problem solving" be added to the list that already includes "critical thinking skills." The member from Ninevah warns the Board that he won't vote for any textbook that says dinosaurs are a proven fact. His motion to that effect fails, but publishers in the room take notice.

As soon as the Board approves the new curriculum and textbook specifications, as amended, 12 textbook company representatives rush the final documents back to headquarters in a hurry because the editorial staffs will have to cull through all the required topics and skills and cross-reference them with the topics and skills demanded by two other major states, Downhome and Serenity. Three smaller markets—Sequoia, Iroquois, Winnebago—are on the same adoption cycle and their needs will have to be considered too.

The editors in the 12 competing companies, however, can't really begin to write the actual text for the new series even though the printer's deadline is only six months away. Before they can begin the creative work, they need to consider the new Curriculum Framework from the State of Downhome, which is behind schedule thanks to political complications. Meanwhile, the State of Serenity has decided on a radically new approach to the curriculum. It is so distinctive that no other state would even consider buying a book based on it.

There is method in Serenity's madness. The State Superintendent not only favors the new approach; she also wants to use her state's economic clout to break up the monolithic textbook system. The Serenity market, she thinks, should be tempting enough to induce some publishers to create an enlightened alternative.

AT THE HARRIMAN COMPANY, there is a high-level meeting to deal with the Serenity crisis. At the outset, the executives announce that the Board of Directors of Humongous Corp., which owns Harriman, has scotched the idea of writing a textbook series just for the Serenity market. "Too risky," they say. "If we don't make it in Serenity, we can't sell the book anywhere else." The new edition will be designed primarily for Nirvana and Downhome, but the editors are also instructed to consider the requirements of 12 other market areas targeted by Harriman.

The sales manager forcefully argues for spending a higher percentage of the project budget on graphic design. He points to the national trend toward all-teacher voting on textbook adoptions. In this "Astrodome method" of textbook selection, he explains, teachers converge on a large gymnasium on the appointed day. They only have time to look at the boxes

and the covers of books, and perhaps a quick flip through its pictures.

"I saw our whole program get shoved to the back of the table because our box wasn't as pretty as our competitors' boxes," he says. Reluctantly, the executives agree to increase the graphics budget and allocate more space in the book to pictures. The editorial director howls with pain because more pictures mean less space for words, but the sales department prevails. "Nobody out there reads the words anyway," says the sales director.

Harriman's executives then discuss the composition of the author team. They agree that Professor Hamilton should be the senior author. Many generations of Nirvana teachers have passed through his classes at State U., and his name on the masthead will enhance sales in that critical market. Hamilton will get 5 percent of net sales. Two teachers, Carmen Hernandez, from Happy Valley School District in the State of Expansion, and Herman Cummings, from Intermediate School #23 in the City of Great Expectations, are also signed on as authors. Both have served as officers in their national teachers associations. Both will receive a modest fee.

Also, the company signs on a long list of consultants from many strategically placed school districts around the nation. Their names will appear on the title page along with brief biographies.

The author team meets together only once at Harriman headquarters, where the project director gives them guidelines on how to proceed. They are instructed to follow the format used by the best-selling series. He passes out copies of the outline, which encompasses the requirements of the 14 target market areas and tells them to follow it.

"Nobody out there reads the words anyway," says the sales director.

They receive a list of words and images to be avoided: junk food, conflict with parents, minorities doing farm work, cantankerous old people, dinosaurs, women wearing the color pink, wine bottles on cafe tables, too much leg, church spires, and flawed American heroes. They are also given a list of items that must be included: kindness to animals, active senior citizens, minorities in professional roles, the work ethic, and flawed American heroes. Each author is expected to write the text for two books in the six-book series.

THE TEACHER guide, which is the most important part of the program, will be done in-house. The peripherals—workbooks, skill sheets, test pack, audio-

tapes, videotapes, and maps—will be farmed out to contractors. Harriman's executives know that adoption committees look with favor on companies that produce a large array of extras. Although most local school districts won't actually buy most of these products, the sales force can offer them as freebies to wavering adoption committees. The costs, of course, can be passed on to buyers in the form of higher prices for the student text.

The project director hires a number of retired teachers and housewives to produce a "Correlational Analysis" for every jurisdiction that demands one. They will cross-reference, by page number, every topic, fact, and skill the jurisdiction wants mentioned or plans to test. In order to get the work done on time, the correlators will have to begin their work before the authors turn in their manuscripts, so inevitably, there will be many errors in the documents. But Harriman's executives know that few administrators or adoption committees actually have time to read the document, let alone check its accuracy. Producing the analysis is expensive and nearly useless, but it has become a business necessity and the cost can be passed on to the buyers.

When the authors' deadline arrives, only Cummings has completed his assignment. He wrote some good lessons and developed some interesting activities, but his prose is jargon-filled and will require extensive editing. Hamilton's mother died and he only had time to prepare an outline. At the last minute, Hernandez, whose district is flat broke, was saddled with three new courses to prepare for and couldn't carve out the time.

When the truck arrives from the warehouse, each member will find it necessary to empty out the garage or the family room in order to house the textbooks.

Caught with only three months to produce camera-ready copy, Harriman's project director executes a contract with a "production shop," whose business it is to rescue textbook publishers in times of crisis. The staff at the production shop is experienced, but there are no subject matter experts on its payroll. In order to make sure the content and pedagogy are consistent with current practice in the field, they must reach out to working teachers in a nearby school district.

These newly hired teachers come to the production shop every afternoon, weary from a full day of teaching. They work late into the night and several weekends to pull together the six manuscripts.

These actual authors, like those whose names will appear as authors, have no previous experience writing textbooks. Although they know a great deal about teaching, they have never written for publication before. The

head of the production shop is thrown into shock when he receives the manuscripts. Their writing is either amateurish or exceeds the allowable readability score. He must hurriedly rewrite what they have submitted, chopping sentences in two and excising interesting words because Harriman threatens to hold up payment until the readability score is lowered.

In the meanwhile, the Nirvana State Board of Education has appointed 17 members to the State Adoption Committee. There is one member from each congressional district in the state. Since service on the State Adoption Committee is seen as a polite form of political patronage, the job is passed around as frequently as possible. According to the state statute, no member is allowed to serve more than once. Since nobody on the adoption committee has ever served on one before, none has any experience with the process or expertise in textbook evaluation.

BACK AT Harriman, the editorial staff must scratch plans to field-test the materials on students because the project is behind schedule. They feverishly rush the copy and artwork to the manufacturer hours before their deadline. The entire creative process—from the first word written to final delivery to the manufacturer—has been accomplished in six weeks.

Just in time, the manufacturer delivers \$500,000 worth of free samples to a warehouse in Mylanta. The warehouse director then ships each member of the State Adoption Committee a complete set of materials being considered for adoption in Nirvana that year. Since all of the members are new to the task, they are completely unprepared for the consequences of being "sampled." When the truck arrives from the warehouse, each member will find it necessary to empty out the garage or the family room in order to house the textbooks and all their accompanying extras. The pile is so daunting that few members dare to touch the books.

At the first meeting, Harvey tells the members what is required of them: they are not to accept calls, gifts, or visits from publishers' representatives; they must complete their work in 60 days so the Board can act on their recommendations in time for the opening of school in September.

The members go home with copies of Helmut's 879 behavioral objectives and Nirvana's standard textbook-evaluation rating sheet with 102 items to be ranked on a scale of 1 to 5. Back in their homes, committee members leaf through the mass of materials aimlessly, not sure of what to look for. Some members alight on pages they don't understand. Some of them conclude that things must have changed quite a bit since they were young, and others conclude that they are too tired to tackle the task and go to bed.

At the next meeting, the committee members begin to fill out their rating sheets. The subcommittee on race and gender proposes the elimination of two books because too many of the females pictured are in passive poses. Although some of the other members secretly think that is a trivial reason, they go along with the subcommittee's proposal because it cuts down the number of books they have to evaluate.

The subcommittee on curricular congruence skims the tables of contents, the chapter heads, the indexes,

and the glossaries. The state agency curriculum specialist, who heads that subcommittee, wants to eliminate one textbook series because the cross-reference to critical-thinking skills in the publisher's explanatory materials only lists 32 references to teaching "main idea," while all the other series have no fewer than 72 references.

The committee has 10 series (60 books) left to consider. Since the main thing is to get the rating sheets filled out, though, the members don't have time to read the books. Of the 17 members, only one has actually read an entire chapter in each of 10 third grade books. The others have actually read no more than three contiguous pages in any of the books. Through many frustrating hours, the members assign high or low marks to "eye appeal," "convenience," "durability," "author credentials," and "content." The members wonder why they are asked to rate "publication date" on a scale of 1 to 5, especially since all of the books bear the same date—next year's.

When Harvey tallies the results, most members are surprised to find that some of the books they liked best didn't even rank in the top five. Harvey prepares a report on the committee's recommendations and sends it to the State Textbook Officer.

MEANWHILE, THE representative of the publisher whose book came in sixth in the rankings is tipped off by a friendly member of the committee. Desperate to get his book on the list, the representative goes to the State Textbook Officer and offers him a deal. If his book can get on the list, his company will give away, entirely free, first and second grade student books for every child in Nirvana. After consulting with his superiors, the Textbook Officer revises the committee's recommendations, putting the sixth-ranked book in fifth place, and removing the fifth-ranked book from the list to be forwarded to the Board for approval. He feels good about what he has done. The state will save a considerable amount of money, and times are hard in Nirvana.

When the committee members find out that the Textbook Officer has tampered with their ratings, most of them are furious. Some members start a rumor campaign about cronyism between state officials and textbook executives, but other members, who couldn't see any remarkable differences between the textbooks anyway, think that the last-minute substitution is a fine idea because it saves money.

At a raucous public hearing before the State Board, a

Some of the best teachers in Nirvana make plans to leave teaching and go into real estate.

colorful array of special pleaders make theatrical presentations favoring or criticizing the books proposed for adoption, and the Board approves the list of books forwarded by the Textbook Officer. Representatives of the winning companies are jubilant, but the losers go home to prepare new resumes.

Two months later, students all over the state open up their beautiful, sweet-smelling new books. A first grader in Chipville opens his book and reads:

"Rabbit said: 'I can run. I can run fast. You can't run fast.' Turtle said, 'Look Rabbit. See the park. You and I will run. . . .'"

Of the 17 members, only one has actually read an entire chapter in each of 10 third grade books.

Soon the child is gazing out the window at the clouds in the Nirvana sky.

A fifth grader in Mylanta reads his assigned chapter and discovers that all the answers to the questions at the end are in the chapter summary. After the first lesson, he learns to skip the chapters entirely, which are boring and confusing anyway, and to read just the summary. He can get rid of homework questions in just a few minutes. "Did you really do your homework that fast?" asks his mother. "Yes, Mom. Now can I watch television?" he asks.

As the year goes by, the teachers in Faraway discover that the kids aren't reading their textbook assignments. A few teachers blame it on "home problems" or television watching, but most of them conclude that the books are too difficult. These teachers form a committee to write to the State Board requesting that readability formulas be lowered in the next adoption.

Some teachers, though, know that their students won't read the books because they are unreadable. They are insulted by the scripts in the teacher edition, which say, "Open the book to page 251, hold it up before the class, and say, 'This map on page 251 shows the nations of Europe.'" Some of the best teachers in Nirvana make plans to leave teaching and go into real estate. □

WHAT SHOULD YOUNG CHILDREN BE DOING? (Continued from page 33)

by the time they are six, give or take a year, are at significant risk in adulthood. There are several kinds of risks. There are risks in terms of mental health, in terms of marital adjustment, parenting competence, and, some say, also occupational adjustment.

Although definitions of social competence vary on some of the details, they generally include the capacity to initiate, develop, and maintain satisfying relationships with others, especially peers. Social competence does not require a child to be a social butterfly. It is not a source of concern if a child chooses to work or play alone, as long as he or she is capable of interacting productively and successfully with another when desired or when appropriate. Research shows that low-accepted and aggressive, rejected children have significant probabilities of being school dropouts. They also have the probability of being delinquents.

The acquisition of social competence involves many complex processes beginning in early infancy. It should be noted that inappropriate, as well as appropriate, social responses are learned through interaction. Weaknesses in social competence may be intensified during such interactions unless adults help the child alter maladaptive patterns. In the preschool period, inadequate peer-interactive skills are unlikely to be improved through formal instruction or even coaching. They can, however, be modified by the intervention of a knowledgeable teacher. Fortunately, a range of techniques that teachers can use to foster the development of social competence is now available.

THE RECURSIVE CYCLE

It is useful to think of social competence as having the characteristics of a recursive cycle. The principle of the recursive cycle is that, once an individual has a given behavior or characteristic, reactions to him or her tend to increase the chances that he or she will display more of that behavior or characteristic. For example, children who are likable, attractive and friendly tend to elicit positive responses in others fairly easily, and because they receive such positive responses, they become more likable, attractive, and friendly. Similarly, children who are unattractive, unfriendly, and difficult to like tend to be avoided or rejected by others. In response to this avoidance and rejection, they tend to behave in ways that make them even more unattractive. This, in turn, increases the likelihood that they will more often be avoided or rejected, and the cycle becomes well established.

The principle of the recursive cycle confirms the point made earlier, that young children should be engaged in interactive processes, especially in the company of teachers who have specialized training and competencies in helping young children maximize the educative potential of interaction. Unfortunately, a young child cannot break a negative cycle alone. Even for adults, breaking a dysfunctional cycle by oneself is very difficult. The young child's capacity to understand the cause of his or her social difficulties and make the

necessary adjustments is virtually nil. Adults must intervene to break faulty cycles. Adults can teach young children more productive peer interactive patterns *in situ* during ongoing social interaction in the early childhood setting.

Recent experience suggests that if we respond to children's needs for help in the development of their social competence in the early years, we can do a great deal to get them on a positive cycle and relieve them of much anguish that inevitably accompanies social difficulties in childhood. If we wait until a child is nine or ten years old and is making life for himself or herself or for others difficult, we may need substantial resources from community agencies to intervene, and still may be too late. These recent insights from research on children's social competence suggest that preschool teachers' concern with social development is well placed and should be given as much weight in planning and teaching as is children's intellectual development.

CURRICULUM OPTIONS

Many people within and without the field of early childhood education think that the choice for curriculum is to have either an academic or a socialization focus. Some of the risks of introducing academic tasks to young children have already been indicated. But the alternative is not simply to provide spontaneous play, though all children up to about seven or eight years of age can probably benefit from spontaneous play. Rather, the data on children's learning seem to suggest that what is required in preschool and kindergarten is an *intellectually* oriented approach in which children interact in small groups as they work together on a variety of projects that help them make sense of their own experience. These projects should also strengthen their dispositions to observe, experiment, inquire, and reconstruct aspects of their environment.

Preschool and kindergarten programs should provide opportunities for interaction; active rather than passive activities; spontaneous play, which can be provided for by setting up the environment with interesting things that children spontaneously enjoy—costumes, sand, blocks, and so on—and group projects that extend over time so that children can strengthen the disposition for sustained interest.

I recently heard indirectly about a group of kindergartners who undertook a detailed study of their own school bus! Though I have no direct knowledge of the entire scope of their project, it is easy to see what kinds of activities the children might have undertaken. One small group could study the driving mechanism, including the motor and gear shifts, brakes, accelerator, steering wheel, and so forth. Another can study the variety of lights inside and outside the bus. It can be noted by children that some lights are for signals, others are to give a warning, and others, of course, are to light the way ahead as well as inside. Other children in the class can study the gauges and dials in the bus and what kinds of information they yield. Another group can take measurements of the width of the bus, count the number of seats, establish how many wheels it has, and learn something about air pressure in the tires. Perhaps two of

them can examine the inside and outside rear-view mirrors. Imagine what kind of vocabulary building can accompany such a study: terms like ignition, emergency door, fuel, dial, gauge, air pressure, accelerator, rear-view mirror, gears, and so forth. Those children able to do so could copy down all the "writing" they could find on the bus and use it for vocabulary and reading studies in the classroom. The door of a school bus, usually opening and closing in accordian fashion, is unlikely to resemble those in the children's homes or their school, so varieties of doors in their immediate environment might become a topic for extended study. Study can extend to the route taken by the bus, who and how many children board at each stop, what traffic signs and signals are passed en route, and so forth.

According to my informant, following the detailed examination of their bus, the children built a bus from scrap materials in their classroom and then acted out a variety of roles associated with transporting children from home to school and back. Virtually all aspects of the work undertaken by the children in this project lent themselves to art work, including drawing, painting, and making plasticene or wooden models.

There is no special virtue to studying a school bus in the sense that some important test will ever examine the knowledge children gained from the project. What is important is that the bus is part of the children's own daily environment and that they learned a lot about it: the correct names of various parts of it, a simple understanding of how it works, and what features of it contribute to their safety. It is especially important that the project provide a context in which children's dispositions to observe, inquire, and become interested and involved in a sustained group effort can be strengthened. In a project of this type, the teacher alerts children to a wide range of potentially interesting aspects of the topic that will take several days or even weeks of continuous probing and exploring. They can ask questions of adults, such as the bus driver and perhaps a mechanic, and look up things in reference books of appropriate levels of difficulty. The fact that the children are expected to tell and explain what they have learned to their own classmates is likely to encourage persistence in attaining information and reaching for adequate understanding. The use of adults, other than teachers, as sources of potentially valuable information can be launched through this kind of project. Furthermore, for many of the children in the class, this project is likely to strengthen the disposition to observe all kinds of other vehicles more closely than they had before, perhaps making useful comparisons and reporting them to their classmates from time to time.

In sum, the project approach can be valuable for young children because it addresses their intellects. It can strengthen a variety of important dispositions, provide rich content for conversation, and offer a context for peer interaction in which cooperative effort makes sense. Projects are also culturally relevant when they stem from the children's own interests and environments. But it should not be overlooked that another virtue of the project approach is that it can make teaching interesting—something very unlikely to be characteristic of more formal academic approaches to early childhood education. □

THE LOSS OF JOBS

(Continued from page 19)

are inadequate; moreover, they can never sustain the necessary political support. What's needed are jobs, created either by the private sector, perhaps thanks to inducements from the government or, if necessary, by the government itself.

He argues that President Johnson's Great Society and War on Poverty were largely ineffective precisely because, unlike Roosevelt's New Deal, they were tied to welfare and thus "failed to relate the fate of poor minorities to the functioning of the modern American economy." Moreover, unlike the New Deal, which "offered a modicum of security for all," the War on Poverty programs were targeted to the poor and ultimately lost the political support of the broader American electorate.

Wilson does not detail his policy prescriptions but offers a general political agenda for rectifying the deteriorating condition of the black underclass based on a long-term policy of full employment. And this agenda can only be achieved through a broad-based political coalition of Americans committed to economic reform and social justice.

"In the final analysis," he writes, "the question of reform is a political one. Accordingly, if the issues are couched in terms of promoting economic security for all Americans, if the essential political message underscores the need for economic and social reform that benefits all groups in society, not just poor minorities, a basis for generating a broad-based political coalition to achieve such reform would be created."

Although there will certainly be those in the civil rights community and among social scientists who will resist Wilson's analysis that today the problems of the underclass are primarily rooted in poverty (class) and not race, there are indications that a number of black political leaders are recognizing what Bayard Rustin and the Rev. Martin Luther King, Jr., realized over 20 years ago—that the plight of this nation's disadvantaged can only be effectively addressed by sound economic and employment policies that will aid all Americans.

Wilson does not pretend to have all the answers. His emphasis on jobs, which clearly are at the heart of the problem, does not, for instance, fully address the question of long-term educational failure and its impact on the ability of ghetto youth to compete in an increasingly specialized and competitive job market. It may be overly optimistic to assume that if jobs were to miraculously spring up in the ghettos, then crime, teenage pregnancy, illegitimacy, and the other social pathologies would disappear. While it is certainly true that unemployment in the ghetto has been the catalyst for changes in cultural norms and behavior patterns, *undoing* these crippling changes may involve not only jobs, but also cultural and educational solutions.

This aside, Wilson has written a truly important and provocative book that should profoundly alter the debate on the origins of the underclass and policy solutions aimed at alleviating its terrible plight. His trenchant insights and recommendations are an invaluable starting point for reevaluating strategies that will ease the suffering and continued social isolation of millions of Americans. □

ARRIVING IN LAKE WOBEGON

(Continued from page 15)

label, but is it really always deleterious? If a teacher or administrator learns from test scores that his students are not up to snuff in long division and therefore puts more emphasis on those skills before the next round of testing, isn't that "teaching to the test" an entirely appropriate use of test results? Indeed, some proponents of test-based accountability explicitly aim to encourage teaching to the test, albeit often under a less disparaging label. "Measurement-driven instruction" is one such label; it refers to the explicit use of tests to push instruction in desired directions.⁹ A currently more popular rubric is "curriculum alignment," which refers to efforts to improve scores by increasing the match between curricula and tests.

The vast gray area of teaching to the test thus stretches from frank cheating at one extreme to appropriate remediation and instruction at the other. Both educators and educational researchers disagree strenuously about where the line between appropriate and inappropriate teaching should be drawn. At what point, for example, does teaching the content specified in curricular goals shade into drill on items that look too much like those that will be on the test? Teachers face dramatically different guidelines in confronting these questions, depending on where they teach. Virginia, for example, provides teachers with guidelines that prohibit, not only teaching actual test items, but also "special reviews or content drills in preparation for testing"; moreover, "instruction or study materials referenced to test results (i.e., focusing on previous test results)" is to be avoided for four weeks before testing.¹⁰ In contrast, other states provide study materials that closely resemble test items. For example, in dismissing disagreement about allowable responses to an item on the Tennessee High School Proficiency Test, an assistant commissioner in that state recently noted that "In the study guide we have *an almost identical item*" [italics added].¹¹ The differences are dramatic at the district level as well. As one district research director said to me wryly, teachers are praised in some districts for behaviors that would earn them censure in his.

Moreover, what is appropriate in one case may be inappropriate in another, depending both on the circumstances in which a test is used and the inferences one draws from it. For example, some educators approve of a brief period familiarizing students who are being tested for the first time with the format of multiple-choice tests and optical-scanning sheets, but

We must remind ourselves that test scores are not synonymous with achievement.

strongly disapprove of spending instructional time on those issues with students who have been tested several times already.

To draw a line between appropriate and inappropriate teaching to the test is beyond the scope of this article, but, for present purposes, it is useful to think of three broad categories. The worst category is instruction that both inflates test scores and degrades instruction—for example, by causing a decline in other skills not measured by the test. The most desirable sort is instruction that produces an increase in scores that is neither spurious nor a signal of instruction that has been in some sense degraded. A middle category is instruction that inflates test scores but that does not actually degrade instruction. The middle and the worst categories shade together, however, for teaching to the test that inflates test scores, if nothing else, wastes instructional time that could be put to better use—a particularly serious cost, given that the United States provides its students with far less instructional time than do most other industrialized nations.

To sort out these different types of teaching to the test, we must remind ourselves that test scores are not synonymous with achievement, even though they are all-too-often treated as if they are. Tests are merely incomplete and fallible indicators of achievement. The inferences about achievement that one can validly make from them depend on many factors, including the characteristics of the test, the characteristics of the instruction students receive, and the way in which the test is administered. Moreover, the value of tests as indicators of achievement can easily be degraded if they are used incorrectly.

In most cases, test scores are valued because they permit us to make inferences about students' mastery of *domains* of achievement that comprise a variety of skills and knowledge, as well as the ability to apply these in a diversity of contexts. A superintendent, for example, might want to tell her board that her students' mastery of basic arithmetic has improved, but it wouldn't be—at least, it shouldn't be—worth a trip to the meeting to announce an increase in facility with subtraction of two-digit numbers, without borrowing, presented in a vertical format. In designing a test, writers must create a set of questions that is a representative sample of the specific skills and knowledge that constitute each domain, so that we can generalize from performance on the test to mastery of the domain.*

If teachers, however, increase the attention they devote to the content of a given test, without a corresponding increase in the attention they give other components of the domains in question, the test

*This is not doing justice to a complex issue. First, there are clear exceptions, although many are largely irrelevant to the issues at hand. For example, if one wanted to ascertain whether a paramedic had mastered cardiopulmonary resuscitation, one might test those specific skills and not think of the results as representing some broader domain of interest. Second, there is considerable disagreement presently about how narrowly domains should be defined, in terms of both measurement and educational considerations. Nonetheless, the characterization here is appropriate for the tests Dr. Cannell described: policymakers and educators take them as indications of mastery of fairly broad domains of achievement.

becomes a less representative sample of the domains. If the increased learning of the content of the test generalizes to the rest of the domains, fine; if not, test scores will be an inflated indication of mastery. That is, the rise in test scores will not signal a correspondingly large increase in the aspects of achievement we hope to measure. A concrete indication that such inflation had occurred would be if students scored substantially lower on a second test that included a substantially different but equally reasonable sample of tasks from each domain. Low scores on the first test would still indicate a lack of mastery, but the meaning of high scores on the first test would be called into question.

***One reading teacher complained,
“We are now spending from
September to November on basic
skills rather than on our
developmental program.”***

Many people apparently believe that this is a purist's concern and that most of the gains on high-stakes multiple-choice tests do indicate corresponding gains in students' mastery of the achievement domains the tests purport to represent. Research, however, indicates that everyone, not just purists, should worry about generalization if teaching is too closely aligned with tests. Research shows that when instruction is tightly aligned with a specific test, scores can be changed markedly by even modest—sometimes, seemingly trivial—changes in the test. For example, in one study, students were taught to convert Roman to Arabic numerals.¹² In the instructional materials, the Roman numerals were always presented first. One test mirrored this aspect of instruction; another was the same except that Arabic numerals were presented first. This trivial difference in test format changed the average score *by 40 percent!* This example, while particularly remarkable, is not unique.

Unfortunately, no one knows to what extent recent gains in test scores are specific to the tests that students have been given. Undoubtedly, some of the gains from accountability-oriented testing are real, while others are spurious. This can be checked empirically, but in most instances no one has bothered to do so. Still, there is more than enough reason to worry about excessive alignment of instruction with the characteristics of specific tests. One warning sign, noted above, is the use in some states of state-provided study guides with material that is “almost identical” to the tests. Another signal is the popularity of commercially produced test-preparation materials, sold both by testing firms and by other educational publishers. In some districts, the use of such materials is encouraged. Some of these test preparation materials are virtually identical to the tests, except for the specific items used. For example, one of

the popular test-preparation books, Random House's *Scoring Higher on the Iowa Test of Basic Skills*, is for all practical purposes just a practice ITBS; even such irrelevant details as the shapes of the op-scan ovals used in different subtests have been accurately mimicked.*

Even apart from the use of explicit test-preparation materials, many teachers end up aligning their instruction in many ways to the specific test used. Interviews with teachers indicate that while many are comfortable with the current emphasis on tests or even welcome it, others are worried and can offer examples of what they consider too much tailoring of instruction to match their jurisdictions' tests. In one study, a teacher complained that “the principal made the teachers take [the test] and rewrite it so it wouldn't be exactly what the students were going to have” to provide preparation. Another argued, “You are more or less forced to teach the format of the test or you come up showing that your students haven't learned what you really feel they have learned.”¹³ One reading teacher complained, “We are now spending from September to November on basic skills rather than on our developmental program.”¹⁴ Lorin Anderson, a researcher who has conducted numerous observational studies of classrooms, has spoken of what he calls the “blanks to bubbles” trend: in some districts, as the time of mandatory testing draws near, fill-in-the-blank worksheets are increasingly replaced by sheets that look much like multiple-choice test items, complete with little bubbles to fill in to indicate the correct answer.

Whether instruction that is tightly aligned to a test teaches the broader skills of interest is a pressing question, for the amount of time devoted to test preparation is in many instances large and growing. Lorin Anderson has found that in one state, he is generally unable to observe instruction for more than a month each spring, because at least that much time is devoted to preparing for and taking tests. Carol Tittle, an educational researcher at the City University of New York, was told by one New York City school that she could not pilot-test a study there any time from mid-January through the end of citywide testing in May because of test preparation. Indeed, in some instances, the line between regular instruction and test preparation has faded from view. A recent *Washington Post* article about a low-income, high-minority school with an average third grade California Achievement Test (CAT) score at the 99th percentile in terms of national norms noted, with no apparent irony, that the school's “formula for success includes . . . a well-organized staff that starts preparing students in kindergarten for the third-grade CAT.”¹⁵

WHAT PRICE ARE WE PAYING?

Dr. Cannell's report provides no clue about the mix of different types of teaching to the test or about the extent to which the inflation of test scores can be attributed to this factor. But the fact that test scores are badly exaggerated in some instances, and the scattered evidence of

*The publication of these test-preparation materials need not indicate the approval or cooperation of the test developers. In this particular case, for example, the director of the Iowa Basic Skills Testing Program, H.D. Hoover, strongly disapproves of materials such as the Random House book.

"I do not do as many essay tests as I did before, because I try to give them things they are apt to meet on standardized tests."

inordinate test preparation, inevitably raises the question: just what are we sacrificing to obtain these inflated estimates of students' performance? How does test preparation compare to the activities that we want students and teachers to be involved in?

One familiar concern is that teaching to the test will narrow the curriculum, making teachers and administrators reluctant to devote time to subject areas not covered by the exams. We know this happens, even if we don't know how much. As one teacher noted, "I do less science. I have always been very strong on science, but you have got to meet the standards of those tests basically in math, reading, and language arts."¹⁶ One administrator complained, "We realize a kid is taken out of science every other day for citizenship and will fail science to maybe pass the citizenship test."¹⁷ As a consequence of this problem, there has been movement in some jurisdictions to expand the range of subjects included in the mandatory testing, but the range is still meager in many places.

Another concern is that the range and depth of teaching may be narrowed *within* subject areas that are tested. This problem may be even more important than insufficient emphasis on certain subject areas, but it is much harder to pin down or to get agreement about what constitutes evidence of it.

One reason for concern about narrowed range and depth of teaching is the types of tests used in most accountability-oriented testing programs. Most rely virtually exclusively on multiple-choice tests.* Multiple-choice tests should not be overly disparaged; there is a great deal that a high-quality multiple-choice test can tell us about students' achievement, and such tests have many advantages—such as consistency of scoring, ease of administration, and cheapness—that make them an almost inevitable component of any testing program. For all their strengths, however, multiple-choice tests have many weaknesses as well. Multiple-choice tests represent a fairly narrow range of tasks. For example, they require students to recognize correctly and incorrectly spelled words, not to spell correctly; to understand the mechanics of language usage, not to write a coherent and well-reasoned passage; to recognize right and wrong answers, rather than to explain why answers differ in their correctness, or to explain what premises

*The one major exception is the writing samples required by many states. Their function, however, is generally limited to appraising students' ability to write; they are usually not used to assess students' mastery in other subject areas.

or viewpoints make a given answer correct.

In part for this reason, some critics argue that success on multiple-choice tests depends substantially on relatively basic skills. Certainly, multiple-choice tests can be designed to place considerable emphasis on higher-order skills. The ACT and the College Board Advanced Placement tests are often noted as examples. There is evidence, however, that as a class, multiple-choice tests are not well suited to assessing certain higher-order skills. Performance on such tests is often determined more by basic skills—such as factual recall—than the designers of the test intend.¹⁸ Some critics also argue that multiple-choice tests are better able to measure components of reasoning than integrated, higher-order thinking.¹⁹ Moreover, many of the tests used in accountability systems are more strongly focused on basic skills than the multiple-choice format demands. This is obvious in the case of many minimum-competency tests but is also true of some other commonly used tests. For example, one study found that two of the more commonly used third grade reading tests emphasized basic skills—in this case, literal comprehension—considerably more than did some common third grade texts.²⁰

The danger, then, is that instruction will focus excessively on the types of tasks represented on the tests, at the expense of others. This would distort, not only our estimates of achievement, but also our efforts to improve it. Indeed, we know that this occurs; we just don't know how much. Surveys of teachers have found that this problem is quite common. As one teacher admitted, "I've changed my teaching behavior [in response to the test] . . . I do not do as many essay tests as I did before, because I try to give them things they are apt to meet on standardized tests. I feel that it is hurting the children, because they don't have to write their own sentences."²¹

These concerns are especially pressing because of the widely acknowledged weaknesses of a large proportion of American students on tasks requiring higher-order skills, such as reasoning, problem solving, and the like. The achievement trends of the past 20 years have, if anything, made this problem worse.²² Many critics of the current wave of testing express concern that an overemphasis on raising scores will lessen the amount of instruction of precisely the sort that would facilitate the growth of higher-order skills. Rather than drilling for such tests, these critics argue, we would do a lot better to have students spend their time working on tasks that do entail higher-order skills: writing persuasive essays, designing and evaluating hypotheses, reading challenging books and analyzing them, and so on.

WAYS OUT OF WOBEGON

Where do we go from here?

To start, we must face the fact that test-based accountability has not always worked as advertised. As Dr. Cannell has pointed out, it has often produced absurd results. In addition, there are disturbing signs that it has substantial unintended costs. In making measurement-driven educational policy a cornerstone of the reform movement, we made a more powerful change in the educational system than many people anticipated. And we made this change with our eyes closed, giving little

credence to evidence that the change was risky and establishing virtually no mechanisms for evaluating or monitoring its effects. Dr. Cannell's report should spur us to open our eyes, to look carefully and systematically at both the good and the bad effects of test-based accountability, and to search for factors that maximize the good while minimizing the bad.

As we search for ways out of Lake Wobegon, we will not find tidy solutions. Rather, we will need to make a long series of compromises between competing goals and purposes. Moreover, we will need fundamental changes in our attitudes concerning the meaning and appropriate uses of tests. Such a reorientation, unlike many technical changes, cannot be effected simply by changes in policy; it will require education, debate, and time.

As we search for ways out of Lake Wobegon, we will not find tidy solutions.

First and foremost, we need to re-establish the basic principle that in most instances, test scores are not an end in themselves. They can serve as a means toward that end—albeit one with many risks—and as an incomplete measure of our success in reaching it, but no more. A rise in test scores can be a very good sign, but it is often insufficient to confirm that the battle has been won, and, absent other indications of improvement, it may be a sign that the battle has been fought on the wrong front.

Second, we need to work out a more reasonable compromise between the legitimate pressure for accountability on the one hand, and the demands of appropriate practice and comprehensive assessment of student progress on the other. Dissatisfaction with American schooling is pervasive and, in many instances, well founded, and the public is not about to give up its freshly asserted right to know whether things are improving. Our tools for establishing accountability, however, are crude. If accountability is to be imposed from afar—from state capitals, for example—there will be unavoidable pressure to use assessment tools that are inexpensive in time and money and consistent in their scoring, such as multiple-choice tests. To obtain a comprehensive view of students' performance, however, and to nudge instruction in some of the desired directions, we need to rely on other assessments that are time consuming and difficult to standardize. Essay tests and term papers, for example, can reveal many critical skills that multiple-choice tests can't discern—such as the ability to piece together a logical argument and to generate and evaluate hypotheses and arguments. Equally important, the process of writing essays and

papers is not merely assessment; it is beneficial instruction in its own right. We can also ask students of the sciences to devise hypotheses and experiments to test them, and judge their ability to do both. We can ask students to debate both sides of an argument and evaluate how well they do so. We can give students two views of a historical phenomenon and see how much evidence they can bring to bear—and with what degree of logic and coherence—to choose one, or to explain why different people have held each. A good teacher could easily expand this list.

How can this compromise be reached? One approach, described below, is to expand the range of tests that can be standardized well enough to be used as accountability measures. Another is more difficult to bring about: we need to avoid letting simpler, accountability-oriented tests become so important that administrators and teachers feel pressure not to use richer forms of assessment in their own instruction.

Third, we need to recognize that when a test is used as a *prod* to raise achievement, its value as an *indicator* of achievement is compromised. Once educators begin worrying about a particular test as an end in itself, it becomes suspect as a sample of the domains of achievement that it purports to measure, for all of the reasons described above. A prudent generalization for interpreting a rise in test scores is therefore that the higher the stakes attached to that particular test, the more one should look for other signs that the rise in scores indicates a real improvement in competence.

Finally, we need to confront the need for more reasonable and consistent standards for appropriate teaching to the test. It is simply indefensible to have teachers subject to dramatically and arbitrarily different standards in this regard, and to have students receive sharply different instructional programs as a result. Both inflated test scores and distorted instruction will persist until this question is addressed.

SOME SPECIFIC PATHS

With these general principles in mind, we can assess a variety of specific approaches that might be used to lessen the Lake Wobegon problem.

Check for Malfeasance. One approach, worthwhile even if it nets only a handful of wrongdoers, is to hunt for malfeasance on the part of both test publishers and educators. Dr. Cannell's calls for external scrutiny of the testing enterprise would be one reasonable step. Another step would be to check test results routinely for suspiciously high scores or unusually rapid rates of improvement. Some states and localities are already doing this—for example, the state of Virginia and Fairfax County, a Virginia suburb of Washington, D.C.

Gather the Evidence. A second approach is to establish reasonable efforts to discern the good and bad effects of different forms of test-based accountability. Currently, debate about its effects is based more on speculation and rhetoric than on evidence. To chart a more rational and productive path in the future, we need to investigate how this sea change in educational policy is affecting what is taught and how it is taught. We

need to monitor the byproducts of the change, such as stress on students and teachers. We need to check whether the gains students are showing on high-stakes tests really generalize—that is, whether their gains in *achievement* are as great as their gains in *scores*. We must evaluate both the costs and the benefits of different approaches.

Several kinds of evidence should be brought to bear in assessing what we have wrought. We need to listen to what the participants—kids, parents, teachers, and administrators—can tell us about their experiences. We need systematic, quantitative investigation of such questions as the extent to which increases in test scores generalize to broader domains of achievement. We need ongoing monitoring to see how schooling changes as a result of test-based accountability. For example, how are decisions about tracking, retention in grade, special education placement, and exclusion from routine testing affected? Are the changes in instruction teachers make in response to high-stakes tests consistent with what policymakers intend?

Improve Testing Programs. A third direction would be to improve testing programs, while making little change in the types of tests used. There are many ways this could be done, and some would have significant benefits. Their value is often debatable, however, because some technical changes that improve testing in some respects have drawbacks as well. One reason is that tests are being asked to serve many different functions—comparing a district to the nation, for example, as well as charting progress over time within a given district—and changes that improve the tests with respect to one of these functions often make them less suited to another.

Annual renorming of standardized tests is an example of a change that would have both advantages and disadvantages.

Annual renorming of standardized tests (one of Dr. Cannell's suggestions) is an example of a change that would have both advantages and disadvantages. On the one hand, it would eliminate the bias from historical happenstance when test scores are compared to a national norm, one of the major causes of the Lake Wobegon effect. Annual norms are not very useful for measuring *progress*, however; a district that is exactly keeping pace with a national rise in scores would appear to be standing still. To gauge progress over time, one needs some reference point that does not change.

Another example would be a switch to "user norms"—that is, norms based on the districts using the test rather than on a national sample of students. User norms solve one major problem in the current system:

they avoid the inflation of scores that occurs merely because the curricula of test users match the test better than the curricula of the norming sample. User norms would exact two prices for this considerable benefit, however. First, unless one has a firm idea of what the other districts that use the test are like, it is hard to know how to interpret scores that are compared to user norms. This would be a serious limitation in the case of tests administered to a large and diverse set of schools and districts. Second, a change in the group of districts using the test—for example, the loss of a few very large districts with particularly high or low scores—could fundamentally alter the meaning of the norms, rendering trends meaningless.

Another suggestion that has received some attention in the wake of Dr. Cannell's report is the possibility of using different versions of the test every year. This would eliminate some outright cheating and might lessen somewhat the most egregious teaching to the test. If the alternate versions greatly resemble each other, however, as they typically do now, one should not expect too great an effect. Even though specific items would be changed, one could still direct teaching toward the *types* of items the test is likely to emphasize.*

As an alternative, one could create a set of tests that are intentionally diverse and rotate among them (in an unannounced order) from year to year. Each test would have roughly the same proportion of its items from each domain specified in the districts' curriculum framework. Certain core skills might be represented in each test. The other test items, however, would be chosen to provide the broadest practical coverage of the domain and a considerable diversity of formats; these would be distributed randomly among the different tests. The breadth of coverage of the set of tests would be greater than that of a single test, and the only way to teach to the test would be to focus on the broader curricular goals of the district. This would avoid both some of the inflation of test scores and some undesirable teaching to the test. This approach has important drawbacks, however. One is cost. Another is that the large differences among the tests would make it difficult to make them fully equivalent. As a consequence, small differences in scores and year-to-year changes would be even less meaningful than they currently are. Large differences, however, would still be meaningful, and *consistent* trends over a period of years would be *more* significant than they are now.†

*Consider, for example, the SAT, which is constantly revised and rescaled. Students and coaching firms nonetheless spend a lot of time practicing the types of items (certain kinds of analogies, for example) that are likely to appear on the test.

†A change of this sort is considerably easier to effect if a test is designed from the outset to assess the achievement of *groups*—schools, say, or districts—rather than that of *individual students*. When scores of individual students are not at issue, each student can be given a small portion of the total set of items, so that the total range of the assessment is far broader than the set of items given to any one student. (This procedure is used, for example, in the National Assessment of Educational Progress to increase the range of material that can be assessed in a relatively brief testing period.) When reliable scores are needed for individual students, however, the range of items given to each cannot be as sharply reduced.

Secure tests might also help, particularly if adults other than district personnel were used as proctors. This too would be expensive, however. Moreover, if tests were to provide a desirable push to instruction, teachers would still need considerable information on the content of the tests. Finding the appropriate amount of information—enough to induce beneficial teaching to the test without inflating scores or degrading instruction—would be a difficult and controversial task.

Another improvement in the use of current tests would be to reform the way scores are presented and used. Testing companies and administrators should be explicit in telling the public what the tests do and don't measure. As much as possible, forms of reporting that exaggerate districts' or states' apparent performance must be avoided. When they cannot be avoided—for example, when comparisons to dated national norms are used—the public should be informed about the factors that might inflate the apparent level of performance. The fact that small differences in scores—just like small differences in public opinion polls—have little or no meaning, either in evaluating students or in ranking schools or districts, should be stressed. Similarly, the public should be reminded that test scores are strongly influenced by many factors that have nothing to do with educational programs. Unfortunately, there is no obvious method for encouraging these changes. Administrators will be tempted to follow this advice when their scores are low and to ignore it otherwise, so careful reporting will often have the flavor of excusing poor performance. For that reason, it may be necessary to undertake a public information effort, designed to better inform the public, policymakers, and educators about the meaning of test scores.

BETTER TESTS, BROADER FOCUS

A fourth direction—which has potentially great payoff—would be to expand the criteria by which we judge education, with the twin goals of providing a fuller view of students' performance and encouraging more beneficial and less undesirable teaching to the test.

One aspect of this approach would be to expand the range of tests that can be used as accountability measures. This may mean creating new types of tests or developing more standardized ways of scoring existing types of tests, such as essays. Dramatic improvement in both of these spheres will be slow in coming, and policymakers should not see this as a quick route out of Lake Wobegon. Nonetheless, a variety of alternative approaches is being developed and tested, and some might offer considerable improvements.

One promising example is a new reading assessment developed by the state of Illinois in conjunction with the Center for the Study of Reading at the University of Illinois. The new assessment, which is intended to be more consistent with research on the acquisition of reading skills, is still a multiple-choice test, but with some important differences. The reading passages are long, for example—typically 900 to 1,000 words, and sometimes longer. The questions are intended to require analysis, and they frequently have more than one right answer. (Interestingly, young children seem to

have no problem with this, but some teachers and high school students, long accustomed to more traditional tests, initially find it unsettling.)²³

One rationale for the new Illinois reading assessment is that it can be successfully "taught to" only by doing what reading teachers should do anyway.

The state of Connecticut has also been breaking new ground, in this case attempting to develop standardized scoring methods for a variety of measures of actual performance. One result is a hands-on component in the state's science assessment, in which science teachers who are trained as raters appraise students' handling of scientific experiments. The state is also adding a spoken-language proficiency test to its foreign-language assessment; again, teachers are being trained to provide standardized evaluations of performance. Similarly, the new secondary education-assessment program in Great Britain, the General Certificate of Secondary Education, includes practical, oral, and classroom-performance components.²⁴ In addition, some long-standing American tests that have been used for other purposes go beyond the multiple-choice format. For example, the College Board Advanced Placement Tests and some of the New York State Regents Examinations include essay components. Careful investigation of the way in which such tests are scored, as well as of their effects on instruction, might provide useful archetypes for improved accountability systems.

Ideally, new approaches such as these, when added to more traditional tests, will offer a more comprehensive and less inflated view of students' performance. In addition, in many instances, their proponents hope that they will lessen undesirable teaching to the test, in that preparation for these assessments will come closer to the forms of instruction that are desirable in their own right. One rationale for the Illinois reading assessment, for example, is that it can be successfully "taught to" only

REFERENCES

by doing what reading teachers should do anyway—for example, by having children read long passages of text and analyze them. Only time will tell, however, whether these efforts have the payoff their sponsors hope for.

One caveat about new types of tests is that some of the more promising new approaches are expensive, in terms of time, money, or both. The Connecticut system, for example, requires that a sizable number of teachers be trained to rate performance, and their ratings take time. The Illinois reading assessment takes more time than a traditional reading test. This is not a problem if it is used only to assess performance levels of whole schools, for in that case, each student need only be given a single passage to read. Evaluating individual students, however, would require more than a single passage, and therefore a considerable increase in testing time. In addition, more complex forms of assessment will, in many instances, have a larger margin of error than some current tests; a lesser degree of precision may be the necessary cost of assessments that come closer to measuring the right things.

For the foreseeable future, however, a truly comprehensive view of students' performance—and of the quality of the schools they attend—will require more than the measures that can be sufficiently standardized to be used for accountability from afar. By one means or another, we must broaden our focus, and we must change the incentives so that teachers and administrators are encouraged to do so as well. Teachers must be encouraged to rely routinely on measures such as essays, term papers, experiments, and debates, far beyond the limited extent to which such measures are standardized for accountability. They must not feel compelled—as many do now—to forsake these richer and more beneficial forms of assessment in favor of classroom assessment that resembles high-stakes multiple-choice tests.

Moreover, we must attend to the aspects of a quality education that will not appear even in these more comprehensive measures of achievement. No set of achievement measures will alone tell us whether students are engaged and motivated, whether they enjoy school, and whether schools are giving them, not just some set of skills and information, but the enjoyment of learning and *intrinsic* motivation to learn that they will need to be well educated and successful throughout their lifetimes.

Making the needed changes in our uses of tests, and refocusing our attention on the broader goals of providing a high-quality and successful education, will be costly. It will require a great deal of debate and work, not just in the short term, but on a continuing basis. It will make both our assessment programs and the answers we get from them more complex. But if the goal of today's greater accountability is really to improve students' learning, we have no alternative.

Dr. Cannell's report is hardly the only evidence that things are amiss. But it has helped alert some people to the fact that the comforting confidence they had in the current, sometimes simplistic approach to test-based accountability was badly misplaced. If we now respond the way our children deserve, the Lake Wobegon report, for all its failings, will have done a tremendous public good. □

¹Georgia to test kindergartners for promotions, *Education Week*, March 2, 1988, p. 1.

²Extensive tests make teachers uneasy, *Education Week*, April 13, 1988, p. 1; *Indiana Statewide Testing for Educational Progress (ISTEP)*, Center for School Assessment, Indiana Department of Education, Spring 1988.

³H. Dickson Corbett and Bruce L. Wilson: Raising the stakes in statewide mandatory minimum competency testing, in R. Crowson and Jane Hannaway (eds.): *The Politics of Reforming School Administration*. Philadelphia, Tyler & Francis, 1988.

⁴H. Dickson Corbett and Bruce Wilson, personal communication, April 20, 1988.

⁵D. Koretz, *Trends in Educational Achievement*. Washington, D.C., Congressional Budget Office, 1986.

⁶Examples can be found in Koretz: *Trends in Educational Achievement*, chapter 2, and R.L. Linn: Testing and instruction: Links and distinctions, *Journal of Educational Measurement*, Summer 1983, pp. 179-189.

⁷L. Dickson Corbett and Bruce Wilson, personal communication, April 20, 1988. This example was uncovered by one of the few systematic studies of the effects of high-stakes achievement testing on instruction. For other results of the study, see B. Wilson and L.D. Corbett: Two state minimum competency testing programs and their effects on curriculum and instruction, in R. Stake (ed.): *Advances in Program Evaluation*, forthcoming.

⁸See Judge Louis M. Welsh (ret.): Desegregation, education, or both? The San Diego desegregation case and the rationale for an order to raise achievement, and Thomas S. Nagel: Desegregation, education, or both? From the inside looking out—findings of an integration analyst (both presented at the annual meeting of the American Educational Research Association, April 25, 1984).

⁹See, for example, W.J. Popham, K.L. Cruse, S.C. Rankin, P.D. Sandifer, and P.L. Williams: Measurement-driven instruction: It's on the road, *Pbi Delta Kappan*, May 1985, pp. 628-634.

¹⁰Virginia Board of Education: *Instructions for Test Preparation and Test Administration, Virginia State Assessment Program (VSAP)*, December 12, 1983.

¹¹Tennessee teacher protests, *Educational Assessment—Policy and Use*, March 1988.

¹²For details on this and other similarly remarkable examples, see S. Alan Cohen: Instructional alignment: Searching for a magic bullet, *Educational Researcher*, November 1987, pp. 16-20.

¹³Linda Darling-Hammond and Arthur E. Wise: Beyond standardization: State standards and school improvement, *The Elementary School Journal*, January 1985, pp. 315-336.

¹⁴Wilson and Corbett, op. cit.

¹⁵P.G. school at head of class, *The Washington Post*, January 14, 1988, pp. B-1, B-3.

¹⁶Darling-Hammond and Wise, *Beyond Standardization*.

¹⁷Wilson and Corbett, op. cit.

¹⁸See, for example, N. Frederiksen: The real test bias: Influences of testing on teaching and learning, *The American Psychologist*, March 1984, pp. 193-202.

¹⁹L.B. Resnick: *Education and Learning to Think*. Washington, D.C., National Academy Press, 1987.

²⁰See Bonnie B. Armbruster, Robert J. Stevens, and Barak Rosenshine: *Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests*, Technical Report no. 26, Urbana-Champaign, Center for the Study of Reading, 1987.

²¹Darling-Hammond and Wise, op. cit.

²²See, for example, Koretz: *Trends in Educational Achievement*; also, National Assessment of Educational Progress. *Changes in Mathematical Achievement, 1973-1978*. Denver, NAEP/Education Commission of the States, 1979; and *Third National Mathematics Assessment: Results, Trends, and Issues*. Denver, NAEP/Education Commission of the States, 1983.

²³Thomas Kerins, Illinois State Department of Education, personal communication, May 1988.

²⁴Department of Education and Science, *Report by HM Inspectors on the General Certificate of Secondary Education: An Interim Report on the Introduction of the New Examination in England and Wales*, Welsh Education Department, 1988.

SAVE UP TO 50% ON MAGAZINES



AMERICAN FEDERATION OF TEACHERS SUBSCRIPTION SERVICES

9 Northern Blvd., Box 258
Greenvale, N.Y. 11548 516-671-7744

*Save up to 50% off the usual subscription rates! *Save up to 80% off the cover prices!
*Order new or renew current subscriptions! *Get hours of reading enjoyment for as little as 33¢ a week!

WHICH PUBLICATIONS INTEREST YOU?

Publication	Usual Price	Your Price
A+	24.97	12.97
American Artist 9 issues	14.00	11.97
American Film	20.00	12.97
American Health	14.95	11.95
American Heritage		24.00
American Photog 8 issues	12.95	5.98
Americana	13.95	9.95
Antique Monthly	18.00	13.00
Art News	32.95	25.95
Artist's Magazine 10 issues	20.00	12.47
Arts & Activities	20.00	17.95
Arts & Antiques	36.00	29.95
Astrology	17.97	8.97
Astronomy 9 issues	18.00	15.75
Atlantic Monthly	14.95	9.95
Audio	19.94	9.97
Audubon Magazine	16.00	15.00
Automobile	18.00	9.00
Autoweek 35 issues		11.50
Backpacker	18.00	9.00
Baseball Digest 10 issues	12.46	7.97
Basketball Digest	9.95	7.97
Bicycling	15.97	12.97
Black Enterprise	15.00	11.95
Bon Appetit	18.00	15.00
Boys Life		13.20
Business Week 52 issues	39.95	27.95
Byte	22.00	19.00
Car & Driver	16.98	11.99
Car Craft	15.94	9.97
Cat Fancy	17.97	12.97
Cats	17.50	12.95
Changing Times	18.00	15.00
Child (Fashion by Taxi)	12.00	10.00
Childrens Digest, ages 7-11	11.95	9.97
Childrens Playmate 4-8	11.95	9.97
Computer!	24.00	14.40
Computer's Gazette	24.00	18.00
Consumers Digest		12.95
Cricket, ages 6-13	19.80	15.75
Cruise Travel	12.00	9.97
Cruising World	19.95	17.95
Discover 8 issues	18.00	15.97
Dog Fancy	17.97	12.97
Down Beat	18.00	9.95
Ebony or Ebony Man	16.00	10.97
Economist 51 issues	98.00	60.00
Electric Company		10.97
Electronic Learning	19.95	16.95
Elery Queen's Mys 10 iss	17.50	9.97
Esquire	17.95	9.95

Publication	Usual Price	Your Price
Essence	12.00	9.00
European Travel & Life	24.00	12.00
Family Home Off. Comp.	19.97	9.99
Family Handyman	9.95	5.95
Fate	15.00	11.95
Field & Stream	15.95	7.97
Fifty Plus	15.00	11.97
Financial World 18 issues	31.00	17.95
Flying	18.98	15.97
Food and Wine	22.00	17.00
Football Digest	12.95	7.97
Forbes 29 issues/year	45.00	29.95
Fortune 27 issues/year	47.97	23.99
Forum	23.00	18.00
Games 6 issues/year		11.97
Gifted Children Monthly		24.00
Golf Digest	19.95	11.98
Golf Illustrated	15.00	7.97
Golf Magazine	15.94	9.97
Gourmet	18.00	15.00
Guns & Ammo	17.94	9.97
Hands on Electronics	28.00	18.95
Harpers Magazine	18.00	11.97
Health 9 issues	16.50	8.97

Publication	Usual Price	Your Price
Mother Jones	24.00	16.00
Motor Trend	19.94	9.97
Motorcyclist	15.94	7.97
Ms Magazine	16.00	10.97
Muppet, ages 8-14		7.50
Musician	18.00	10.97
The Nation 24 issues	14.30	9.95
National Law Journal	59.00	29.50
New Republic	56.00	28.00
New Woman	15.00	11.97
New York 50 issues	35.00	19.98
2 years		39.96
N.Y. Review of Books	34.00	31.95
New Yorker 52 issues	32.00	20.00
2 years		40.00
Newsweek 52 issues	41.00	21.80
2 years		43.60
Odyssey, ages 8-14 9 issues	12.50	11.00
Old House Journal	16.00	13.95
Omni	24.00	15.96
1001 Home Ideas 8 issues	14.65	11.97
Organic Gardening	12.97	9.97
Outdoor Life	13.94	8.97
Outside	16.00	8.97
Ovation	21.00	11.95
Parenting	18.00	12.00
Parents	20.00	12.95
PC Computing	19.94	14.97
PC 22 issues/year	34.97	21.97
PC World	29.90	17.97
Penthouse	36.00	30.00
Personal Computing	18.00	11.97
People 52 issues	61.88	30.94
Petersen's Photographic	15.94	7.97
Playboy	24.00	19.00



Our price for 52 issues of PEOPLE is only \$30.94! Compare this price to other offers and calculate your savings!

Just 60¢ a copy

High Fidelity	13.95	6.98
High Technology	30.00	24.00
Hitchcock Mys 10 iss	15.00	9.97
Home		15.00
Home Mechanix Illus	11.94	8.97
The Homeowner	14.40	11.97
Home Viewer	18.00	9.97
Hot Rod	17.94	9.97
Humpty Dumpty ages 4-7	11.95	9.97
Inc	24.00	12.00
InCider	24.97	21.97
Inside Sports	18.00	11.97
Insight	17.00	12.75
Instructor	20.00	12.97
Jack & Jill ages 6-8	11.95	9.97
Jerusalem Post 23 issues	20.75	15.97
Jet 52 issues	36.00	26.00
Ladies Home Journal	19.95	11.97
Lakeland Boating	15.94	11.94
Learning '88	18.00	9.00
Life	31.96	16.25
McCalls	13.95	6.98
Mac World	30.00	24.00
Metropolitan Home	18.00	11.97
Modern Electronics	16.97	12.97
Modern Photography	13.98	6.99
Money	31.95	15.99



Only 56¢ a copy

Through our program you can order 52 issues of TIME for only \$29.12... a \$72.28 savings off the newsstand price!

Playgirl	20.00	17.50
Popular Photography	11.97	5.99
Popular Science	13.94	8.97
Practical Homeowner	12.97	9.97
Prevention		13.97
Radio Electronics	17.97	15.97
Road & Track	19.94	12.99
Rolling Stone	19.95	15.95
Runner's World	19.95	12.97
Sail	21.75	17.95
Sailing World	21.75	10.88
Salt Water Sportsman	19.95	16.95
Saturday Evening Post		12.95
Science Digest 6 iss/yr	12.95	9.95
Scientific American	24.00	19.97
Self	15.00	10.00

This is our **LAST AD THIS SCHOOL YEAR** So order now!

Publication	Usual Price	Your Price
Seventeen		15.95
Shape		20.00
Ski Magazine	9.97	6.97
Skiing	11.94	5.97
Skin Diver	19.94	11.94
Sport	12.00	7.97
Sporting News 55 issues	59.95	29.95
Sports Illustrated 54 issues	64.26	32.13
2 years		64.26
Stereo Review	11.97	5.99
Sylvia Porter's Per. Fin.	19.97	11.97
Teaching and Computers	19.95	16.95
Teaching Pre K-8	15.00	12.98
Tennis	17.94	8.97
Theatre Crafts	30.00	24.95
3-2-1 Contact		12.97
Time 52 issues	58.25	29.12
2 years		58.25
Tours & Resorts	12.00	9.97
Travel & Leisure	29.00	15.00
True Story	15.00	9.97
Turtle pre-school	11.95	9.97
TV Guide 52 issues	39.00	33.80
Unique Homes 6 issues	29.97	19.97
US Magazine	23.95	14.97
US News 52 issues	34.50	17.25
USA Today 260 issues	130.00	97.50
Vanity Fair	12.00	7.80
Vegetarian Times	24.95	19.95
Video	12.00	8.97
Video Review	12.00	7.97
Village Voice 52 issues	37.44	19.95
Weight Watchers	13.97	11.97
Women's Sports & Fitness		12.95
Workbasket or Workbench	6.00	5.00
Working Mother	11.95	6.48
Working Woman	18.00	15.00
World Press Review	24.97	16.95
Writer's Digest 10 issues	17.50	12.47
Yachting	19.98	12.97
Yankee (New England)	18.00	14.95
YM	14.00	10.95

PLUS HUNDREDS OF OTHER TITLES



**AFT SUBSCRIPTION SERVICES, Box 258
9 Northern Blvd., Greenvale, N.Y. 11548**

To save up to 50% on your magazine subscriptions, please fill out and mail in this coupon.

NAME _____

ADDRESS _____

CITY _____ STATE _____ ZIP _____

YOUR SCHOOL _____

PUBLICATION NAME	YEARS	PRICE
1. _____	_____	_____
2. _____	_____	_____
3. _____	_____	_____
4. _____	_____	_____
		TOTAL _____

Check enclosed (made payable to AFTSS) or
 Visa or _____ Exp. _____
 M/C# _____ Date _____

All subscriptions are for one year unless otherwise noted.
New orders: Publishers take from 8 to 12 weeks to start a subscription.
Renewals: Please send us the address label from your magazine at least 8 weeks before the expiration date.

S688

Co-star with Jack Lemmon in a Union, Yes! commercial.

Join the "Why I said Union, Yes!" campaign. Let America know what your union has done for you. Show people the positive side of our unions. And help workers who don't have a union to think "Union, Yes!" If your entry is selected, you'll co-star on a national commercial with Jack Lemmon.



Tyne Daly (Cagney and Lacey) and Howard Hesseman (Head of the Class) are spreading the "Union, Yes!" message. They're the first two of many stars helping our unions get the message home on national television—starting May 11th.

You'll be seen on network TV this fall. Your commercial will be part of labor's next high visibility network TV schedule this fall.

Yes! I want to co-star in a "Union, Yes!" commercial. In 25 words or less, tell how your union helped you solve a problem on the job. Send your name, address, telephone number, name and number of your union local, and the job you perform to: "Why I Said Union, Yes!" c/o the AFL-CIO, P.O. Box 27543, Washington, D.C. 20006. Your entry must be postmarked no later than July 4, 1988.

Employees and officers (and their families) of the AFL-CIO, its affiliates, and agencies are not eligible.



American Federation of Teachers
555 New Jersey Ave., N.W.
Washington, DC 20001

Nonprofit Org.
U.S. Postage
PAID
Permit #4144
Southern Maryland
Division
20790