
Do America's Schools Need a 'Dow Jones Index'?

Although public schooling is often the focus of America's measurement mania, we have yet to devise a comprehensive and useful indicator of the state of U.S. education. Mr. Guthrie argues for the creation of a national education index and suggests the specific items it should incorporate.

.....
By JAMES W. GUTHRIE

IT IS DIFFICULT to envision a people more preoccupied with performance than Americans. Keeping records, shattering records, breaking world records, setting national records, establishing personal records, or being the first, the best, or the most are all the "stuff" of our national obsession with measuring individual and institutional performance. Almost every American city, whether it be a metropolis or a hamlet, lays some claim to a record. It somehow possesses the largest, oldest, longest, heaviest, slowest, tallest, greatest, smallest, tastiest, deepest, quietest, fastest, highest, or prettiest something. It is little wonder that the *Guinness Book of Records* is regularly among the best-selling publications in the United States. (We know because we keep records.)

JAMES W. GUTHRIE is a professor of education at the University of California, Berkeley, and co-director of Policy Analysis for California Education (PACE). This article was initially commissioned by the National Center for Education Statistics, U.S. Department of Education, for the congressionally authorized Special Study Panel on Education Indicators. The author wishes to express his appreciation to John Ralph, John Evans, and Allan Odden for their constructive suggestions. The views expressed are those of the author.

People go so far as to invent activities so that, even if only for a short time, they can hold the record for doing it.

Some of this measurement is frivolous — such as a national survey conducted in 1990 to determine, on the basis of Roloids sales, which U.S. cities had the highest rates of heartburn. Some is fundamental to human survival, such as the figures we maintain on global warming or on infant mortality. Some measurement is straightforward and easily understood, such as annual rainfall records. Other measures are abstract, esoteric, and highly specialized, such as Federal Reserve money supply indicators, the M series. Some measurement is remarkably precise, such as lifetime major league baseball batting averages. Other measures are continuously controversial and subject to constant revision, such as international indices of civil liberties. Some measures are easily calculated and popularly understood, such as won/lost figures for athletic teams. Other measures, while perhaps widely accepted, are only vaguely understood by laypersons, such as the Consumer Price Index.

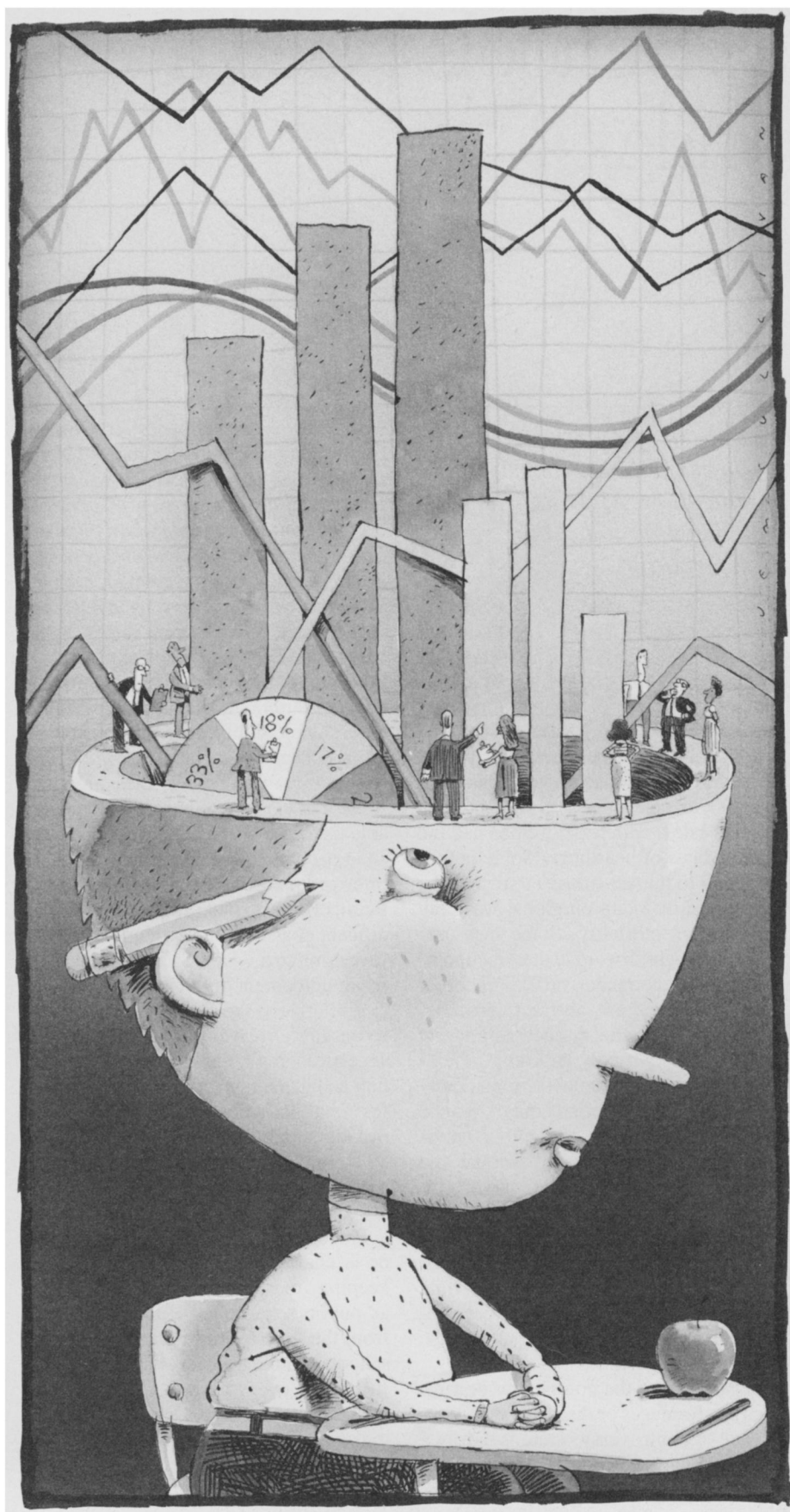
Regardless of the complexity or simplicity, advantages or disadvantages, confidence or controversy, there is hardly a nook or cranny of everyday American existence that goes unmeasured: life, death,

sex, taxes, crime, athletics, economics, transportation, health, commerce, and so on. We as a people have grown accustomed to a broad spectrum of performance measures, which appear regularly on our television, in our newspapers, and in our conversations and which even become part of our everyday contractual agreements, such as pay raises, home loans, and divorce settlements.

No doubt the ultimate explanation for this measurement mania resides somewhere deep within our national character or collective psyche. Meanwhile, suffice it to say that these measures also serve many practical purposes. They enable us to chart trends in areas that have an impact on everyday existence, to make informed predictions regarding important future events, to plot progress toward significant goals, to convey complicated information to a wide audience quickly, to reach agreement on controversial issues in a relatively short period of time, and so on. In short, Americans find performance measures to be a major asset in plotting and planning our personal lives, professional activities, public policies, and private sector endeavors.

MEASURING EDUCATION

Probably no other public sector endeavor is characterized by as much record keeping, measurement, and assessment as public schooling.¹ Records are kept from the time a child is first enrolled in school until the time he or she graduates. Just about every aspect of education — including teachers and textbooks, buildings and budgets, taxes and tests, assets and attitudes — is systematically measured and the data recorded. Moreover,

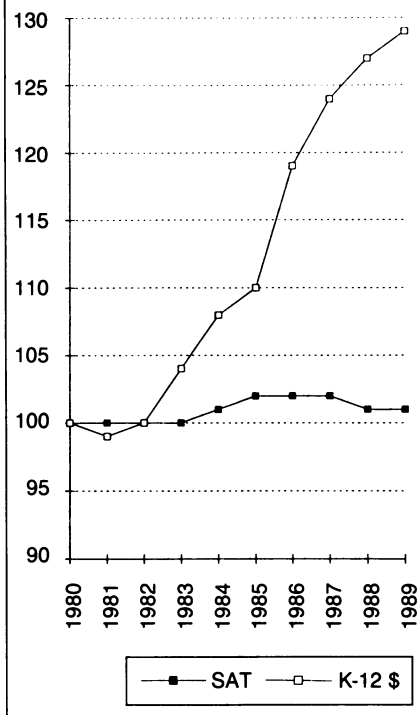


the measurement is multifaceted. Individual schools, local school districts, county agencies, state departments of education, federal government agencies, professional associations, numerous research organizations, and private sector companies are engaged in gathering and compiling these measurements.²

However, simply because educational measurement occurs on a broad scale does not mean that the efforts are accurate, understandable, or useful. Figure 1, which is an adaptation of a graph that recently appeared in a national newspaper, illustrates the point. The graph presents figures in a way that strongly suggests that public schools are inefficient. As resources (spending) have increased, productivity (student achievement) has remained the same.

The graph displays a historical comparison of what is said to be total U.S. spending on education with nationwide student achievement. We are not told how

FIGURE 1.
An Example of Inappropriate Use of Data
School Spending and Student Achievement: Relative Changes In SAT Scores and Total U.S. K-12 Spending, 1980-89



school spending is measured – whether it includes capital outlay as well as recurring expenses. We are not told that spending figures are to be read as billions of dollars. We do not know if this graph includes nonpublic as well as public elementary and secondary education. The spending figure is not adjusted on a per-pupil basis. (Enrollments have been increasing since 1983. It is possible for aggregate spending to increase while per-pupil spending stays even or declines.) We are not told if the expenditures are adjusted for inflation.

However, regardless of the problems on the expenditure side, the measurement of student achievement is distorted enough to constitute something close to fraud. The graph depends exclusively on the Scholastic Aptitude Test (SAT) as an indicator of student performance. The SAT, an indicator of student achievement widely accepted by the media, is an examination developed by the Educational Testing Service (ETS) under contract to the College Entrance Examination Board (CEEB). The latter is a chartered consortium of higher education institutions that collaborate regarding the design and conduct of admissions procedures.

Despite repeated disclaimers by ETS, CEEB, and almost every testing expert in the known world, the public and the press continue to regard the SAT as a measure of what schools teach or students learn. In fact, the examination is designed only to predict freshman success in college. It is analogous to a test of physical agility. Such an examination might measure one's prospective success as a basketball player but would be quite different from an assessment of whether someone could actually play basketball. Aptitude and achievement are related, but they are not the same thing.

Moreover, SAT questions are not now, nor have they ever been, designed to measure what secondary schools teach. The examination is not linked systematically to the curriculum offerings of secondary schools. Rather, it is carefully honed by a cadre of extraordinarily talented technicians to have predictive validity. Test questions that are easily answered by all test-takers are generally eliminated. Similarly, only a few exceedingly difficult questions are included. The remaining test items are included because of their ability to predict college perform-

ance during the freshman year. No effort is made to link the tests to what is contained in the high school curriculum or textbooks – or to what teachers or school boards believe is significant. There is not even an effort to link test questions to what colleges and universities believe is important for students to have learned in secondary school. In short, the SAT does not measure student achievement.

There are other substantial difficulties with the graph, aside from the inappropriate measure of student achievement. For example, only high school students interested in attending college typically take the SAT. The SAT provides no measure of non-college-bound students who do not take the examination, nor does it separate out community college students, who often do take the examination. What are we to assume regarding their achievement? Do they know more or less than was the case in some past period?

Furthermore, not every university uses the SAT to measure prospective success in college. Major universities in approximately half of the states rely on a different examination, one published by the American Council on Testing (ACT). Now, what are we to suppose – that states where SAT scores are stable or declining are the ones in which school spending has been increasing? Or is it perhaps the case that SAT-using states have actually had stable or declining school spending and ACT states have experienced increased school spending?

Answers to questions such as these might or might not make a big difference for the thesis being portrayed by the graph. However, because the media – and thus the public – are accustomed to an oversimplified and inaccurate achievement measure, we cannot tell. Measurement is not automatically accurate, even if printed in a reputable newspaper.

Flaws in the graph raise larger questions. If not SAT scores, then what should policy makers and the general public use as an appropriate measure of educational productivity? Dropout rates? However important school persistence may be, surely it is not by itself a sufficient indicator of the education system's success.

Then what about college admission? This indicator would capture both secondary school persistence and academic

achievement. Or would it? Is admission as a freshman to a community college the same thing as being admitted to Stanford or Cal Tech? What about the large percentage of youngsters who graduate from secondary school but immediately enter the work force or join the military? Would they then show up in a college admissions indicator scheme as a negative mark against a high school?

The National Assessment of Educational Progress (NAEP), initiated for the nation as a whole in 1966 and currently being retooled to permit appraisals of student achievement on a state-by-state basis, will eventually provide at least a partial solution to the problem of measuring productivity. The NAEP will produce student achievement scores for the nation and for individual states in subject areas such as reading, writing, mathematics, and science. However, this measure is, by itself, also likely to prove insufficient. It will not, for example, tell us how many youngsters drop out of school, how many perform well at work, or how many return to college later in life.

In short, it is difficult to imagine any single indicator that could capture the complexities of schooling accurately enough to stand by itself as a measure of educational "productivity." Similarly, the measurement of other dimensions of schooling – for example, "resource inputs," student characteristics, and schooling processes – would entail equally complicated issues. Rather than try to rely on single indicators, advocates of sophisticated school measurement propose the use of so-called composite indicators.

COMPOSITE INDICATORS

The Consumer Price Index (CPI), Gross Domestic Product (GDP), Air Quality Index, and NFL Quarterback Performance Index are all examples of composite indicators. They rely on information gleaned from separate dimensions and combined into a single number. Of course, it is important that the measured components have a fundamental affinity or functional integrity. The measurement nodes for the foregoing indices, while separate, are related to an underlying theme – cost of living, economic production, air quality, or quarterback performance.

The number resulting from the summing of separate measures may be large

(the Dow Jones Index now hovers around 3,000). It may even be artificial (the measure for quarterbacks enables them to perform at higher than 100). The "number" may be the result of substantial mathematical manipulation in order to assign appropriate weighting to sub-components or to convert varying sub-indices to a common and understandable metric. Whatever the underlying statistical procedures, the final "number" is technically justified, professionally defended, analytically employed, widely displayed, and generally accepted. However, it is not necessarily understood by the public.

It would be a rare layperson, selected at random from a city sidewalk, who could even remotely specify the names of the corporations whose daily stock prices are compiled to form the Dow Jones Industrial Average.³ Similarly, not many individuals, other than the technicians who are intimately involved with it, are familiar with the components of the market basket of goods and services that is regularly used by the Bureau of Labor Statistics in compiling the Consumer Price Index. The same could be said for air quality indices that are issued for cities and regions. Probably not even many economists, not to mention laypersons, could clearly explain the manner in which the Department of Commerce produces quarterly estimates of the GDP.

The list could continue, but the point would be the same. The fact that the general citizenry is uninformed regarding the technical bases of these sophisticated composite indicators is no apparent barrier to their professional use or widespread public acceptance.

There is hardly a significant component of Americans' lives for which there is not an applicable, regularly issued composite performance measurement or status metric. Our national economy, collective health, physical environment, personal relationships, and leisure and recreational activities are all relentlessly recorded, compiled, and portrayed by batteries of composite indicators.

COMPOSITE INDICATORS FOR EDUCATION?

Education appears to be the only major societal activity that lacks a publicly accepted composite indicator. But what

might composite education indicators look like? What would they indicate? What areas or educational activities might they reflect? Let me suggest some important topical areas to be considered and provide an illustration of how a performance indicator might operate.

To the question of what dimensions of education should be incorporated into an education index, student performance is an obvious answer and will be addressed below. However, what more, if anything, should be included? At least three additional candidates appear worthy of discussion: public support for education, the conditions of children, and the quality of educational service.

Public support for education. Professional educators contend that, if measuring school or student performance is a good idea, then it is only fair to measure *public support* in return.⁴ How might the degree of public support be determined? Data might be gathered on such items as the results of annual public opinion polling regarding school performance,⁵ expenditures for schooling as a percentage of personal income,⁶ voter turnout for school board elections, the mean teacher salary as a percentage of national mean personal income, and the views of college freshmen on teaching as an occupation.

Conditions of children. Advocates of compiling data on the conditions of children contend that the nation needs to know not only about schooling and student performance but also about the overall status of children. In this view, the U.S. needs to see young people who are 18 and under as a significant national resource on which the nation's long-term well-being will depend. Therefore, an index that incorporated measures of children's health, mental health, criminality, poverty, family stability, personal attitudes, and so on would be a valuable barometer of the nation's investment — or of its need to invest — in the development of this resource.

Quality of educational service. Some educators believe that it would be to the advantage of the public and policy makers to be able to refer to another composite indicator that would convey the availability and quality of a variety of educational services. Into this index could be fitted measures of the upkeep and adequacy of school buildings (e.g., avail-

ability of science laboratories), the availability of advanced placement courses, the quality of teaching, length of the school day and school year, availability of Head Start programs, extent of library or information resources, and teacher/pupil ratios.

ADVANTAGES AND DISADVANTAGES

Is there really more to be gained than lost from having one or more national composite indicators in education? What would be the advantages and disadvantages of having a national education index?

Advocates believe that there are definite advantages to adopting one or more composite indicators of education nationally. They contend that the design and deployment of such measures would both create positive outcomes and eliminate several nagging negative conditions that now exist in the absence of widely accepted composite indicators.

Critics and those who are dubious about the utility of composite education indicators maintain that some things are better left the way they are — "if it ain't broke, don't fix it." Their principal concern is that one or more composite indicators might give both producers and consumers of education an oversimplified picture and encourage schools to pay undue attention to what is measured and insufficient attention to other significant matters.

Advantages. Proponents of an education index say that it would perform a variety of useful functions.

- **Monitoring progress.** If appropriately constructed, a composite education indicator or a set of such indicators could provide a baseline against which to judge U.S. progress in education. Such an indicator could simultaneously serve as a thermometer of the present and a basis for establishing realistic targets for the future.

- **Fostering accountability.** If we assume that what is to be measured is important for schools to accomplish, then a composite indicator would heighten educators' sensitivity to the need to be productive. This would be an even more likely outcome if the indicator were developed so that state, district, and school measures could be compared to national measures.

- **Facilitating communication.** A properly constructed composite indicator need not be completely understood by the general public to be useful. It does have to have sufficient technical validity to win the endorsement of professional experts. However, assuming such a threshold condition exists, then public understanding can be fostered. The indicator would serve to compress complicated information into an easily digested format. If the index included several measures of knowledge or skills, then it might dampen the public's current tendency to become obsessed with inappropriate measures, such as the SAT.

- **Promoting awareness.** An indicator's visibility would promote greater public awareness of education. If, for example, education indicators were issued quarterly, as are CPI and GDP figures, they would be used repeatedly by the media, and the public would develop a heightened consciousness of the status of American education.

- **Enhancing support.** Advocates of an education index contend that the public believes education is important for the nation's long-term well-being and would be willing to allocate more re-

sources to the education system if there were a means for easily conveying its performance and processes. The news need not always be good in order to justify more money. Presumably, low performance measures might also buttress requests for added resources. Regardless of the direction of the results, however, a convenient communication mechanism would enhance the prospect of added resources by engaging the public more effectively in the debate.

Disadvantages. Critics and skeptics contend that widespread reliance on a select few composite education indicators would vastly oversimplify the purposes and processes of schooling. They raise the following specific objections.

- **Inappropriateness.** Education is an intensely personal endeavor. There is as yet no strict science of instruction. Teaching and learning continue to have embedded in them much that is unknown and, to some degree, magical.

Critics of composite indicators sometimes claim that education is too important to be reduced to a single number. To subject schooling to scientific scrutiny and measurement might eviscerate its human core. The arguments are philosoph-

ical and impassioned. The inspiration that a book can trigger, the personal elation at mastering a scientific or mathematical principle, or the fundamental positive influence that schooling exerts on a youngster's self-esteem is simply too important to try to compress into a number that purports to portray the spectrum of goals expected of schools. To have such a number is to make schools the equivalent of widget factories, and that should not happen.

The ultimate objective of schooling is to instill in students a desire to learn. If this is accomplished, then they will assume a larger responsibility for their own education. The act of measurement, particularly if it relies heavily or exclusively upon quantification, seems philosophically inimical to the humane and spiritual essence of true education, critics of composite indicators contend. To subject schooling to an excessive amount of measurement and appraisal is to squeeze the vitality from the process and to intervene inappropriately in the complicated and delicate relationship between student and teacher.

- **Distortions.** In addition to the harmful wedge that measurement can drive between teacher and learner, a heavy reliance on indicators, particularly composite indicators, can distort the purposes and processes of educational institutions.

The assumption, at least among those employed in the institutions subjected to measurement, is that, if a dimension or an activity justifies measurement, it must somehow be important. Given this assumption, it is also important to determine how to enhance the agent's or agency's "score" on whatever is being measured. Activity is then directed — perhaps disproportionately — toward what is to be measured and deflected away from other — perhaps more important — activities. The more important the consequences of measurement are perceived to be, the greater the likelihood that the measurement process will occupy the minds of those engaged in the activity.

The distortion may be of several types. For example, the measured activity may replace other activities, even more significant ones. This is known among organizational sociologists as "goal displacement." Another possibility is that those believing themselves to be judged or rated by the measure may attempt to



elevate scores by illegitimate means — for example, by manipulating the measure, by inducing students to cheat, by offering rewards for student performance, or by actually altering measured results. Any system that encourages dishonest or unprofessional behavior should be questioned.

- **Cost.** The direct cost involved in collecting, compiling, and distributing the information connected with a composite indicator, or even a set of such indicators, is unlikely to be high when compared with the aggregate costs of American education. The endeavor would cost \$5 million to \$10 million at the most. Next to the \$400 billion the U.S. annually spends on formal education, this does not seem like a particularly high price — especially if composite indicators contributed positively in the manner described above under “advantages.”

It is the indirect costs associated with data collection — the time taken away from teaching and learning — that might be high. If the composite indicator could be based on data that were already being collected, then the amount of time involved would be a minimal consideration. However, if producing a composite indicator required yet another intrusion into the world of principals and teachers, then their time would once again be drawn from their chief undertaking, educating students, and spent on a secondary purpose.⁷

- **Misleading.** Critics assert that numbers simply do not convey a sufficient amount of the “truth.” Unemployment figures do not capture the discomfort and loss of dignity among unemployed individuals and poverty-stricken families. The Consumer Price Index may display relatively low inflation for the entire nation, but if any component of the index — such as energy prices — is particularly important to a specific region, then the index does not accurately reflect that region’s economic circumstances.

The same could be true of an overall education index. While it might display student achievement as rising, the achievement of particular groups of students might be stable or declining. The nation’s performance could be rising, but that of a region could be falling. The average for the composite indicator might be rising, but the performance of students in the lowest-achieving quartile might ac-

tually be declining. Critics contend that a composite indicator could not begin to capture and reflect this kind of complexity and thus would run the risk of misleading the public and policy makers.

A composite education indicator could also be misleading if it concentrated exclusively on academic achievement. What about other schooling outcomes, such as creativity, citizenship, compassion, cultural awareness, and ethics? If these important dimensions of education are not measured, won’t they be neglected by schools?

- **Premature.** Critics contend that education is technically too immature to justify the use of complicated statistical indices of output or process. For example, a composite indicator will, of necessity, require the weighting of various component measures. However, with regard to education there is currently little empirical knowledge of what the weights should be. In a student performance index, how should dropout rates be weighted in relation to scores on achievement tests? Answers to such questions rely heavily on human judgment and lightly on science. Yet the overwhelming impression that composite indicators convey to the public is one of technical accuracy and scientific validity.

COMPOSITE indicators have good and bad features. We all live with the day-to-day distortions inherent in the Consumer Price Index, in the Index of Leading Economic Indicators, and in statistics regarding the balance of trade. A decision has been made that, regardless of their weaknesses, these composite measures assist more than they handicap policy makers, economists, planners, investors, and the general public. Information on the economy, the environment, health, and crime is crucial for the nation and for individuals. So is information on education.

The time has clearly come to invest in the development of a mega-indicator or a set of mega-indicators for American education. Arguments against an education index are remarkably similar to those made 50 years ago against collecting unemployment and consumer price information. A single number will distort reality. Technically such data have never

been compiled and distilled, and the supporting knowledge base is insufficient. The public will not understand such a complex concept. Policy makers will not use the information correctly, and so on.

In a small sense, critics of a mega-indicator for education are correct. Initial measurement efforts and any new metric will be imperfect. Sharpening and switching will undoubtedly have to take place. However, education has now gained a permanent place of high priority on the nation’s agenda. The time has come to develop the mega-indicators that the endeavor deserves. After all, with the carefully honed efforts of researchers, statisticians, methodologists, and policy analysts, we can certainly derive a better measure of education than the SAT.

1. A study of children in California found that data on education were far more available than data on juvenile crime, foster care, or children’s health and welfare. See *The Conditions of Children in California* (Berkeley: Policy Analysis for California Education, 1989). Of course, not everyone believes that all available education data are useful. See, for example, Janet A. Weiss and Judith E. Gruber, “The Managed Irrelevance of Federal Education Statistics,” in William Alonzo and Paul Starr, eds., *The Politics of Numbers* (New York: Russell Sage Foundation, 1987), pp. 363-91.

2. Apparently no one has ever assigned a dollar figure to these local, state, and federal efforts to gather education data, but they must surely be expensive.

3. Ironically, the Dow does not measure what many members of the public think it does. It is widely perceived to measure the spectrum of financial market activity. In fact it is intended to measure only industrial stock activities. It omits bonds, commodity futures, metals, services, and a host of other financial market activities.

4. This position is also justified in light of America’s longtime ambivalence about intellect and education. On the one hand, high school graduation and college degrees are viewed as something of a pathway to secular salvation and economic security. School persistence is encouraged, and dropping out is said to be bad. On the other hand, “ivory tower” and “egghead” are disparaging descriptors, “Yankee ingenuity” and the experience of “meeting a payroll” are often thought to be more useful than book learning, and the popular media frequently play up stories of success in the face of academic adversity.

5. Such a poll is now conducted annually by the Gallup Organization in cooperation with Phi Delta Kappa.

6. This figure would be adjusted for fluctuations in student enrollment.

7. Assuming conservatively that each survey or data collection effort for the composite indicator required one hour each from a principal and a teacher from a sample of 4,000 schools four times a year, the total number of hours spent would be 32,000. If the assumption is made that the teachers’ and principals’ time is worth roughly \$40 per hour, then the costs would amount to less than \$1 million. However, 32,000 hours is a lot of time that might have benefited students. K