The National Assessment's Most Useful Background Items[1]

Herbert J. Walberg

University of Illinois at Chicago and
Stanford University Hoover Institution[2]

In 1867, Congress passed legislation to create the U.S. Office of Education (now the Department of Education), which was chiefly to collect educational statistics with a view toward improving education in the nation. Given the No Child Left Behind of 2001 and other federal, state, and local efforts to raise achievement, it is more urgent than ever to collect accurate, comprehensive data for raising achievement, educational productivity and helping to increase the quality of national life.

In keeping with the Department's and the National Assessment Governing Board's ("NAGB") first and perhaps still most important priority, the National Assessment of Educational Progress ("NAEP") should obviously measure achievement. What else? Background information that is required for reporting and information on conditions and policies that may be causally related to achievement, the subject of this paper.

The No Child Left behind Act requires reporting by disability and limited-English proficiency in addition to race, ethnicity, gender, and socioeconomic status. At the request of NAGB, this paper chiefly considers the other category of proven, probable, and controversial causes of achievement for which background or supplementary information would be desirable.

As a founding member of NAGB and chair of the first chair of the Design and Analysis Committee, my long-standing view is that NAEP itself should adamantly avoid causal assertions. Nearly all "X causes Y" statements based on NAEP are likely to be controversial if for no other reason that scholars in various disciplines differ on the rules for proving or probing causality. Many psychologists, for example, view randomized experiments as the "gold standard" of causality, but economists, political scientists, and sociologists most often derive causal inferences from statistically-controlled survey data (oddly, some purely education researchers appear to prefer anecdotes).

For NAGB to assert or even suggest causation might jeopardize NAEP's primary mission and excellent reputation for providing reliable information on achievement levels over the years and for participating states. Mistaken it would be to risk NAEP's reputation, especially at a time when the new federal legislation places an even larger

---

[1] This paper is for the sole use of the National Assessment Governing Board and is not intended for any other group or individual to publish, quote, reproduce, or cite since it draws upon on-going research by the author and does not include emphases, caveats and possible revisions to be discussed at a NAGB-sponsored meeting in September 2002.

[2] University Scholar and Emeritus Research Professor of Education and Psychology, and Distinguished Visiting Fellow.

responsibility on NAGB and as the states become more closely monitored and held accountable for raising achievement.[3]

Moreover, the NAGB board members, and the staff, consulting experts, and contractors that work on NAEP know much about sampling, assessment design, psychometrics, and closely related matters. But they are not necessarily experts about drawing causal inferences, and it seems best for them to avoid "mission creep."

Others, however, should be free to draw causal inferences from NAEP and dispute them in the spirit of free inquiry. Supported for about three decades at considerable public expense and constituting a valuable national asset, NAEP is arguably both the largest and longest continuing education survey ever undertaken anywhere. Since NAEP data are largely under the Freedom of Information Act except for certain identifying information, and since they have been collected at public expense, they should be available to all citizens.

For those who want to provide useful analyses for improving education policy and practice, valid causal inference is the first criterion. If they have no clues about what works, of what value is the research? Scholars, of course, must be open to criticism and discussions about their assertions, and are subject to peer review in refereed journals, scholarly reviews of their books, and criticism of the their work by their institutional colleagues and by others at scholarly meetings. Recognizing that scholars have long carried out causal analyses of NAEP data,[4] the question taken up here is, what data would be most valuable to them?

---

[3] NAEP's and NAGB's reputation are hard won. Getting valid data from surveys is no little matter and requires great care. "Lies, damn lies, and statistics" said Benjamin Disraeli and Mark Twain. The British statistician Sir Josiah Charles Stamp (1880-1941) warned:

> "The government are (sic, British usage) very keen on amassing statistics. They collect them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damm pleases"

Quoted in William Alonso and Paul Starr, (1985) "A Nation of Number Watchers," Wilson Quarterly, 1985, 9 (3), 93-123.

[4] About 20 years ago, I served as principal investigator of a National Science Foundation-sponsored project that first made NAEP data available to scholars and that founded the still active Special Interest Group "Research Using NAEP Data" of the American Educational Research Association, the major education scholarly organization. Other special interest groups since formed indicate the continuing scholarly interest in such analysis include Advanced Studies of National Databases, Large-Scale Assessment, Multiple Linear Regression, Rausch Measurement, School Indicators and Profiles, and Structural Equation Modeling. Scholars not active in the AERA, particularly economists and policy analysts, also make use of NAEP data.

What Suggests Causality?

This paper draws on (1) meta-analyses (statistical analyses of results of many studies) of control-group research and (2) large-scale surveys that seem likely to reveal the causes of achievement. These two kinds of research complement one another. Psychologists prefer control-group experiments, particularly those that randomly assign students to educational methods and conditions, and measure achievement before and after to assess progress. Such experiments have causal creditability, because differences in learning are attributable only to differences between experimental conditions and the luck of the draw, just as in the case of medical experiments that randomly assign patients to alternative regimens.

Experiments, however, are usually weak in generalizability since they typically use small and possibly atypical samples of students, such as those in a given urban or suburban school. Meta-analysis or the statistical analysis of many experimental control-group studies, however, can compensate for the weakness of any single study, since the pervasiveness of an effect can be ascertained by statistically analyzing a variety (usually all) of samples. Such analyses can show whether an educational method works for a variety of students, conditions, and subjects such as for boys as in urban, suburban, and rural students at various grade levels and in various school subjects.

Analyses of state, national, and international surveys can also reveal the generalizability of findings. Using regression and related analyses, epidemiologists, economists, political scientists, and sociologists conduct such research, which usually encompasses whole populations or random samples. These analyses, however, yield less certain causal inferences, since they statistically "control" for alternative causes. In achievement research, these usually include socioeconomic status and other factors, which are usually poorly specified and measured. Such analyses of large-scale surveys may omit plausible causes, since measures of them were left out of the surveys originally designed for purposes other than the analyst's. Since the apparent results may depend heavily on controversial initial assumptions, different analysts of the same data may arrive at different conclusions.[5]

In the last decade, survey analyses improved remarkably, particularly in measuring learning rather than achievement, that is, assessing "value-added" gains or learning over, say, the year from a pretest to a posttest. Analyzing achievement at a single point in time may be misleading, since achievement may be attributable to prior causes, such as infant poverty or prior good or bad teaching, rather than to current conditions or methods.

Though economic, sociological, and political factors affect learning, their influence is indirect. Learning is fundamentally a psychological process; student motivation, instruction, and other psychological factors are the well-established, consistent, and proximal causes of learning. Thus, this paper begins with psychological factors before analyzing the social conditions that affect learning less directly.

---

[5] NAEP's lack of individual achievement measures on more than one occasion is its biggest limitation for drawing causal inferences particularly about effects on individual students.

Psychological Productivity Factors

Table 1 shows three sets of nine factors derived from an early meta-analysis of 2,575 study comparisons that suggested that these factors are the chief psychological causes of academic achievement (and, more broadly, school-related cognitive, affective, and behavioral learning).[6] Subsequent meta-analyses showed results consistent with the original findings. [7]

Evidentiary Basis

Perhaps more relevant for NAEP are findings from large-scale statistically controlled analyses of NAEP data. Several investigators gathered the results of empirical tests of the nine-factor productivity model. Borger (1983) compiled the results of analyses of nine surveys of 15,802 13- and 17-year old students tested in mathematics, science, social studies, and reading by the National Assessment of Educational Progress (NAEP; see also Walberg, 1986). The correlations of the factors with achievement and subject-matter interest as learning outcomes varied from -0.45 to +0.68; 83 or 91% of the 91 correlations were in the hypothesized directions. All factors are hypothesized as positively affecting achievement and attitudes except for amount of exposure to leisure-time television, which is expected to be negatively signed. When the factors were controlled for one another in multiple regression, 58 (or 91%) of the 64 coefficients were signed as hypothesized. [8]

Paschal and Stariha (1989) compiled the results of further regression studies of other data sets including the following U.S. national samples: NAEP, High School and Beyond, the Scholastic Achievement Test, the School Health Educational Evaluation, and the Study of Mathematically Precocious Youth. Included also was the large survey study of mathematics of primary and secondary students in 12 countries completed by the International Association for the Evaluation of Educational Achievement. Several local surveys were included: Chicago preschoolers, Brazilian 12- year-olds, and community college students. Paschal and Stariha combined their results with Borger's to summarize

---

[6] Herbert J. Walberg, "Improving the Productivity of America's Schools," *Educational Leadership* 41, no. 8 (1984): 19–27. Subsequent larger collections of research syntheses are discussed elsewhere including Barry J. Fraser, Herbert J. Walberg, Wayne W. Welch, and John A Hattie, "Synthesis of Educational Productivity Research," *International Journal of Educational Research* 11 (1987): 73–145; and Herbert J. Walberg and Jin-Shei Lai, "Meta-Analytic Effects for Policy" in *Handbook of Educational Policy*, Gregory J. Cizek, ed. (San Diego, CA.: Academic Press, 1999), 419–453. See Margaret C. Wang, Geneva D. Haertel, and Herbert J. Walberg, "Toward a Knowledge Base for School Learning," *Review of Educational Research* 63 (1993): 249–294.

[7] The studies included randomized experiments, and statistically controlled quasi-experiments and well measured but well measured correlational studies. The experimental and quasi-experimental studies were concentrated on the quality and amount of instruction, home environment, and peer environments.

[8] Jeanne B. Borger, Thomas Weinstein, and Herbert J. Walberg, "Synthesis of Multivariate Productivity Itudies." Paper presented at the American Educational Research Association, 1983.

the results of 23 studies of about 250,000 students in six curriculum subjects. Of the 341 regression weights, 303 or 89% were in the hypothesized directions.[9]

How the Factors Appear to Cause Achievement

Each of the first five factors—prior achievement, [10]development, motivation, and the quantity and quality of instruction—seems necessary for learning in school. Without at least a small amount of each, the student may learn little. Large amounts of instruction and high degrees of ability, for example, may count for little if students are unmotivated or instruction is unsuitable. For this reason, each of the first five factors appears necessary but insufficient by itself for effective learning.

These five essential factors, however, are only partly alterable by educators since, for example, the curriculum in terms of lengths of time devoted to various subjects and activities is partly determined by diverse economic, political, and social forces. Ability and motivation, moreover, are influenced by parents, by prior learning, and the students themselves. Thus, educators are unlikely to raise achievement substantially by their own efforts alone.

Of the remaining factors--the psychological climate of the classroom group; enduring affection and academic stimulation from adults at home; and an out-of-school peer group with learning interests, goals, and activities--influence learning in two ways: Students learn from them directly; and these factors indirectly benefit learning by raising student ability, motivation, and responsiveness to instruction. In addition, about ten (not the more typical 30) weekly hours of television viewing seem optimal for learning, perhaps because more television time displaces homework and other educationally constructive activities outside school.

The major causal influences flow from aptitudes, instruction, and the psychological environment to learning. In addition, however, these factors also influence one another, and are also influenced in turn by how much students learn, since those who begin well learn faster.

Thus, in my view, school and district economic, political, and sociological characteristics and conditions may be attenuated influences on achievement because they are less alterable, direct, and observable. They are not substitutes for the nine factors, but more distant forces that can support or interfere with them.

---

[9] Roseanne A. Paschal and Winnie Stariha, "Educational Productivity Studies: A Quantitative Synthesis," in H.C. Waxman (1990) *Study of Learning Environments Monographs: volume 5*. Perth, Australia: Curtin University Centre on Science Education. The earlier studies employed ordinary least-squares regression on cross-sectional samples; some later studies employed longitudinal data and analyses that better take into account measurement error and indirect and multi-level effects.

[10] Achievement may have a large apparent effect because a pretest given a year before a posttest may embody the prior learning previously caused by the other factors, a good reason for having longitudinal data for statistical control.

More and less productive classes, moreover, may be expected in the same school; and it is somewhat misleading to characterize a whole school or district as effective--just as it is less accurate to characterize an optimal condition of plant growth as the average annual rate of rainfall in a state or farm than the amount of rain and irrigation that reaches the roots of a single plant in a given time period.

The educational productivity theory itself is admittedly over-simplified because learning is clearly affected by school and district characteristics as well as many economic, sociological, and political forces at the school, community, state, and national levels. Yet these characteristics and forces--such as the sex, ethnicity, and socioeconomic status of the student, the size and expenditure levels of schools and districts, and their political and sociological organization--are less alterable in a democratic, pluralistic society; are less consistently and powerfully linked to learning; and appear to operate mainly through the nine factors in the determination of achievement, even though, of course, they may be worthwhile measuring and analyzing.

Many of the nine factors are reasonably well represented on past surveys. Sample survey items and descriptions of sets of items from NAEP, High School and Beyond, and the International Association for the Evaluation of Educational Achievement provide national and international baselines in various years, and they should be considered as candidates on this ground alone. Better items and direct observations can also be developed and used.

The discussion now turns more specifically to promising specific indicators of the factors, starting with perhaps the most important for educational practices, the quality and amount of instruction.

Quality of Instruction

Table 2 shows the effects of instructional methods divided into nine categories. These can be more broadly grouped: graphic representation, especially in the form of road maps of what is to be learned; goal setting; and feedback provide direction and redirection. Identifying similarities and differences, summarizing, and generating and testing hypotheses require students to think and express ideas in forms different from presentations. Cooperative learning provides opportunities for students to assimilate and present ideas by explaining various aspects to one another. Homework and practice are indexes of engaged study time. Reinforcement and recognition provide incentives for performance.[11]

The largest collection of estimated effect sizes, which covered 275 methods and conditions,[12] provides further illustration. Discussed here are several of the largest effects

---

[11] The effects reported are based on research in which investigators generally insured implementation. In practice, the various methods need to be reasonably well implemented to insure similar effects. On the other hand, long-term, well-managed implementation might result in larger effects.

[12] Herbert J. Walberg and Jin-Shei Lai, “Meta-Analytic Effects for Policy” in *Handbook of Educational Policy*, Gregory J. Cizek, ed. (San Diego, CA.: Academic Press,

from that collection, including those for traditional methods that have large effects, several newly published effect estimates, and a few selected to show the range of inquiry about instructional quality (see Table 3), on which the following discussion is based.[13]

General Methods

The Elements of Instruction can be considered the most fundamental psychological variables in learning. Cues present what is to be learned and how to learn it. Engagement is the degree to which learners actively participate. Corrective Feedback signals mistakes and furnishes redirection. Reinforcement—one of the largest general effects uncovered—provides encouragement and information that learning is correct.

Mastery Learning combines the Elements of Instruction and requires mastery of learning units before students proceed to the next unit of instruction. In particular, it allows some students as much as five times more instructional time and additional Cues, Corrective Feedback, and Reinforcement. Computer-Assisted Instruction can provide these elements to each student individually. Though beneficial to students in general, even college students, it appears particularly effective in developing skills among handicapped students and those in the early grades.

Direct Instruction can be viewed as traditional or conventional whole-group teaching done well. Specifically, it consists of phases: (1) daily review, homework check, and, if necessary, reteaching; (2) rapid presentation of new content and skills in small steps; (3) guided student practice with close monitoring by teachers; (4) corrective feedback and instructional reinforcement; (5) independent practice in seatwork and homework with high, more than 90 percent, success rates; and (6) weekly and monthly reviews. Comprehension Instruction is similar and consists of three phases: (1) modeling, in which the teacher exhibits the desired behavior; (2) guided practice, where the students perform with help from the teachers; and (3) application, in which the student performs independently.[14]

1999), 419–453. These estimates are based on control-group rather than correlational research and mostly include experimental studies that randomly assign students or classes to groups or quasi-experiments that roughly equate the groups by measuring gains from a pretest before enacting the method to a posttest afterwards, or employing statistical methods such as covariance or regression of determining the “net effect” or “value-added learning gain” after adjusting for initial achievement and other variables.

[13] To my knowledge, the effects are the best but hardly infallible estimates. They depend, for example, on the quality of the underlying research; still many studies by many scholars in many circumstances and pointing in the same direction compel more creditability than any single study, no matter how well planned and executed. The ways of synthesizing research have improved during the past quarter century, but the newer ways usually yield similar results. So older studies are not necessarily excluded here, since they may be the only ones available.

[14] An interesting variant is “reciprocal teaching,” in which paired students take turns leading dialogues on pertinent features of a text. By assuming the planning and executive control ordinarily exercised by teachers, students learn planning, structuring, and self-management similarly to the way tutors learn from teaching, and they learn why it is said that if you want to learn something well, teach it. Comprehension Instruction encourages

Most of the other results in Table 3 can be broadly summarized under the rubrics of the Instructional Elements. Goal Setting, Adjunct Questions, Explanatory Graphics, and Frequent Testing provide Cues, Reinforcement, and Corrective Feedback. Homework, especially with comments and grades, provides Engagement.

Special Methods

Some instructional methods, though they exemplify general principles discussed in previous sections, apply only to particular skills. Consider reading, perhaps the most important skill learned before and during schooling. Phonemic Awareness, Repeated Oral Reading, and Phonics provide beginning readers with mastery of sound-and letter correspondences they may not have learned at home, in preschool, or in kindergarten.[15] Writing may be best taught by writing practice, that is, having students express in the own words what they have inquired about. They can also learn by applying questions and criteria such as clarity and concision to their own and others' writing and then making improvements. Combining their own sentences with those of others adds to their skill in employing appropriate sentence variety.

Grouping allows increases in instructional suitability. Accelerating gifted students allows them to learn at a faster pace without detracting from other students' learning. Tutoring tailors instructional elements to each student. Mainstreaming "handicapped" students into regular classes rather than segregating them in all-day or "pull-out" programs avoids stereotyping and stigmatizing them and helps them make normal progress. What they usually need is more and better, not special, instruction.

The last set of results in Table 3 shows that teachers themselves benefit from Instructional Elements, particularly feedback on their classroom practices, whether on new methods of teaching or on those that should be in their repertoire. New learning of difficult teaching skills may require specific practice with Cues, Reinforcement, Corrective Feedback, and Engagement until they reach Mastery, just as in the case of students.[16]

Early Childhood Programs

Because children in poverty often failed to thrive in the early grades and fell increasingly behind in the later grades, Head Start and other preschool programs have been provided for the last three decades. A 1985 meta-analysis of about 300 studies of these programs revealed that their moderate immediate effects on achievement and other cognitive tests faded within two to three years; that is, program students did better on achievement tests than control-group students at the end of the program, but the

---

students to measure independently their progress toward explicit goals—a big lesson in life.

[15] National Reading Panel, The *National Reading Panel Report: Teaching Children to Read* (Washington, DC: U.S. Department of Health and Human Services, Public Health Service, National Institute of Health, National Institute of Child Health and Human Development, 2000).

[16] If both teacher- and self-instruction are considered, perhaps these elements apply to much of human le.arning, including such diverse fields as sports, ballet, chess, music, foreign languages, and the professions.

difference between the groups diminished to insignificance.17 During the intervening period, the programs appeared to improve by concentrating on children's academic readiness, and recent reviews are more encouraging.[18]

The only long-term study of an academically focused school-related program showed significant long-term effects and cost-effectiveness. The Chicago Child–Parent Centers (CPC) provided academic and family-support services to children, beginning at age 3. The program emphasized the acquisition of language and premathematical experiences through teacher-directed, whole-class instruction, small-group activities, and field trips. Parental participation in the program was intensive, with coordinating activities in each center's parent resource room.

Compared with matched control-group children, the 989 CPC children in the program showed higher cognitive skills at the beginning and end of kindergarten, and they maintained greater school achievement through the later grades. As reported in the *Journal of the American Medical Association*, a study funded by the National Institutes of Health and the Department of Education, showed that, by age 20, CPC graduates had substantially lower rates of special-education placement and grade retention than the control group, a 29 percent higher rate of school completion, and a 33 percent lower rate of juvenile arrest. A cost–benefit analysis showed that at a per-child program cost of $6,730 for 18 months of part-day services, the age-21 benefits per child totaled $47,759 in increased economic well-being and reduced expenditures for remediation. [19] Very few education studies have either followed children as long or calculated the costs and benefits of the programs.

Amount of Instruction

An extensive meta-analysis of correlations and effect sizes showed that 88 percent of the 376 were positive.[20] For intervention studies, the average effect-size for more time was .47 which compares favorably with effects of good teaching methods. The research

---

[17] Karl R. White, "Efficacy of Early Intervention," *Journal of Special Education* 19 (1985): 401–416.

[18] Lynn A. Karoly, Peter W. Greenwood, and S. S. Everingham, *Investing in Our Children: What We Know and Don't Know about the Costs and Benefits of Early Childhood Interventions* (Santa Monica, CA: RAND Corporation, 1998); Janet Currie, "Early Childhood Programs," *Journal of Economic Perspectives* 15 (2001), 213–238.

[19] Arthur J. Reynolds, *Success in Early Intervention: The Chicago Child-Parent Centers* (Lincoln, NE: University of Nebraska Press, 2000); Arthur J. Reynolds, Judy A. Temple, Dylan L. Robertson, and Emily L. Mann, "Long-Term Effects of an Early Childhood Intervention on Educational Achievement and Juvenile Arrest: A 15-year Follow-up of Low-Income Children in Public Schools," *Journal of the American Medical Association* 285 (2001), 2339–2346.

[20] Herbert J. Walberg, "Uncompetitive American Schools: Causes and Cures." In Diane Ravitch (Ed.), *Brookings Papers on Education Policy* Washington, DC, Brookings Institutions, 1998, pp. 173-226.

suggests beneficial learning pay-offs of a longer school year, longer school hours, long-term study including homework, regular attendance, intensifying instruction by efficient time use during instruction, and matching time allocated to time needed. Good teachers appear relatively adept at efficient time use, even though large amounts of classroom time can be wasted partly because usual teaching methods fail to suit low- and high-ability students. Studies of economically disadvantaged, handicapped, and other categorized students show positive benefits of time and greater needs for both extending and intensifying learning time to reach a given standards.

Student Aptitude

Since such things as socioeconomic status and motivation cannot easily and directly be changed, psychologists and sociologists have long studied achievement correlations with measures of student backgrounds. Even substantial and consistent correlations, however, may be weak indicators of causality, since they usually lack experimental or statistical controls. They are, nonetheless, worth measuring and considering just as consistent correlations of cigarette smoking and lung cancer hardly prove but do suggest consideration of causality, particularly if they corroborate other evidence, say, causality in experimental studies of mammals exposed to varying degrees of cigarette smoke. Measures of student aptitude and background are also valuable in statistically controlling for other policy and practice factors. In addition, some student characteristics such as motivation can be indirectly altered by incentives, as indicated by both experimental and multivariate studies.

Prior Knowledge

Table 4 expresses correlations as effect sizes comparable to those in Tables 2 and 3. Students' prior knowledge has a huge predictive effect, perhaps since knowledgeable students can increase their learning from a bigger base. Previous success may also motivate students..

Motivation

Motivation itself is closely associated with how much students learn. Multivariate analysis of surveys and control-group studies of reinforcement corroborate its causal influence. Perhaps the most exciting demonstration of motivational effects is the Dallas O'Donnell Foundation Advanced Placement Incentive Program. The Foundation paid both teachers and students $100 for each Advanced Placement examination passed.

In the nine participating Dallas public schools, sharply increasing numbers of boys and girls of all major ethnic groups took and passed the AP exams. The number rose more than twelve fold from 41 the year before the program began to 521 when it ended in 1994–95. After terminating, the program continued to have carryover effects: In the 1996–97 school year, two years after the program ended, 442 students passed, about 11 times more than the number in the year before the program began.[21] This massive effect

---

[21] Herbert J. Walberg, "Incentivized School Standards Work," *Education Week*, 4 November 1998, 48. Many education theorists deny the role of incentives and hold that true or superior learning only takes place when it is intrinsically valuable to the student. But there is little evidence that students are unaffected by long- and short-term external

sharply contradicts the prevalent idea in education that learning must be intrinsically motivated for its own sake.

Home Environment

The effect of the Home Environment can be taken very seriously for several reasons. Control-group studies corroborate many correlational findings. The Home effect is far larger than apparent socioeconomic effects. Something can be done about Home Environments: School–parent programs can help parents academically stimulate their children by reading to them, taking them to libraries, guiding and discussing leisure television viewing, cooperating with home visitors and teachers, talking with their teachers, becoming informed about school, and similar practices. These may also be indicated by the presence of objects in the home such as dictionaries, encyclopedias, books, and computers.

Socioeconomic Status

By comparison to home environment influences, measured socioeconomic ("SES") effects are small. They may only be weak indicators of the actual home environment that directly affects students' academic development. Children may have only a hazy idea of their parents' income, occupation, and education. Because of divorce, separation, and other family changes, the parent that student questionnaires ask about may not be the parent of preceding years that exerted long, strong influences on the child's development. Besides, parents can change in socioeconomic status during the first 18 years of students' lives making a current indicator misleading.

Aggregate Effects

School-, district-, state-level research is generally less rigorous than studies of individual children and classes. Why? If half the teachers in a school practiced ineffective methods and half practiced effective methods, the net result would be an average teaching effect, which would conceal important effects within the school. Many school-level studies, moreover, have inadequately measured and controlled for prior achievement and other productivity factors with strong records of affecting learning. Even so, for the sake of completeness, the possible school-level influences are worth considering; particularly those corroborated by control-group research and statistically controlled analyses of student and classroom effects.

Curriculum Alignment

Table 5 shows a strong influence of Opportunity to Learn, which refers to the extent that education goals, curriculum, instruction, and testing are "aligned." Most centrally, this means that what is tested overlaps with what is taught.[22] Aside from the Australia,

---

incentives. Even if they were unaffected, they need preparation for adult life, and most adult work, with the notable exceptions of that in bureaucracies and public schools, employs merit pay, that is, rewards results.

[22] Opportunity to Learn results comport with common sense. Students taught Japanese would undoubtedly obtain better reading and listening test scores than students not taught Japanese.

Canada, Germany, and the U.S., most nations have national education systems, which allow such alignment across schools in each country. Many individual states such as North Carolina and Texas are aligning their systems of education, so that if education goals are X, Y, and Z, curricula, teaching, and testing are geared not toward M, N, and O but toward X, Y, and Z.

Goal Setting

Psychological studies support the idea of setting national, state, and local achievement goals. Laboratory control-group research and field studies in a wide variety of organizations confirm the effects of setting goals on task performance. Nearly all studies showed that setting specific, challenging goals led to higher performance than setting easy goals, "do your best" goals, or no goals. "Goals," it has been concluded,

> affect performance by directing attention, mobilizing effort, increasing persistence, and motivating strategy development. Goal setting is most likely to improve task performance when the goals are specific and sufficiently challenging . . . feedback is provided . . . the experimenter or manager is supportive, and assigned goals are accepted by the individual.[23]

Other School Effects

Table 5 shows that School-Level Instructional Time, Student Monitoring, and Parental Involvement influences are positive and coincide with classroom- and student-level research. The school-level effects are smaller, however, perhaps because, as noted above, they average important differences among classrooms and students within schools and because they may be unreliably reported on questionnaires rather than observed. Perhaps because they are vague and difficult to measure, School Climate, Administrator Leadership, and Staff Cooperation are the weakest apparent school-level influences.

Possible Policy Influences

For two reasons, NAGB might consider meaning possible policy influences. Several possible influences appear consistent significant in statistically controlled survey research. Others are hotly controversial, and require more research. My provisional candidates for a list of possible indicators for NAEP measurement in Table 6 are discussed briefly in this section.

External Examinations

The Cornell economist John Bishop intensively studied effects of curriculum-based external examination effects on learning. He analyzed surveys of the examination effects on learning of the (U.S.) Advanced Placement program, the New York State Regents, and U.S. state and Canadian provincial systems. He also analyzed examination effects on learning in the United States in comparison with effects in Asian and European nations. The examinations have the common elements of being externally composed and geared

---

[23] E. A. Lock, K. N. Shaw, L. M Saari, and G. P. Latham, "Goal Setting and Task Performance" *Psychological Bulletin* 90 (1981): 125–152. Quote on p. 125.

toward agreed-upon subject matter students are to learn within a nation, state, or province. Often given at the end of related courses, they have substantial positive effects on learning.24 Made publicly available, the examinations allow citizens, policymakers, educators, parents, and students to assess and compare achievement standings and progress.

Accountability

A decade ago, few states specified what students should know and be able to do, but 49 states now do so, and the number of states with adequate academic standards has increased. The more sustained and comprehensive the accountability system, moreover, the better states' learning progress appears. A study commissioned by the National Educational Goals Panel revealed the reasons that North Carolina and Texas made the largest gains on the National Assessment of Educational Progress:

- grade-by-grade standards with aligned curricula and textbooks,
- expectations that all students would meet the standards,
- statewide assessments linked to the standards,
- accountability for results with rewards and sanctions for performance,
- deregulation and increased flexibility in ways the standards can be met, and
- computerized feedback systems and achievement data for continuous improvement.[25]

Policy analysts have begun rating the states for both standards and accountability, which to be most effective, must presumably go together. Good standards are rigorous, clear, written in plain English, communicate what is expected of students, and can be assessed. Good accountability systems are aligned with the standards and include school report cards, ratings of schools, rewards for successful schools, authority to reconstitute failing schools (for example, by replacing the staff), and the actual exercise of such legislated consequences. Only five states—Alabama, California, North Carolina, South Carolina, and Texas—have solid standards and strong accountability systems. [26]

---

[24] For a summary, see John H. Bishop, "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning," *International Journal of Educational Research* 23 (1996): 653–752.

[25] The authors also attributed the gains in the two states to the intensity and stability of business support for the reforms but not to per-pupil spending, pupil/teacher ratios, proportion of teachers with advanced degrees, and average of teacher experience. See David Grissmer and Ann Flanagan, *Exploring Rapid Achievement Gains in North Carolina and Texas* (Washington, DC: National Educational Goals Panel, 1998) and also the authors' "Searching for Indirect Evidence for the Effects of Statewide Reforms" in *Brookings Papers on Education Policy 2001*, ed. Diane Ravitch (Washington, DC: Brookings Institution, 2001), 181–229. See this volume for other papers finding constructive achievement effects of accountability.

[26] Chester E. Finn and Marci Kanstoroom, "State Academic Standards," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2001*, ed. Diane Ravitch (Washington, DC: Brookings Institution, 2001), 131–180.

Small Schools and Small Districts

In the half century through 1990, the number of U.S. school students rose from 25.4 million to 41.2 million. The number of districts, however, declined from 119.9 thousand to 15.4 thousand, and the number of schools declined from 247.1 thousand to 81.7 thousand. During the period, accordingly, the average number of students per school rose from an average of 103 to 504, and the number of students per district rose from 214 to 2,683. The distribution of both schools and districts is positively skewed; there are a few huge ones and many relatively smaller ones concentrated in rural areas particularly near the Canadian border.

The massive increases in school and district size took place despite research showing that large organizations tend to become departmentalized, impersonal, bureaucratic, inefficient, and lacking in individual and institutional accountability. Their goals tend to become diffuse, and they tend to be more subject to needs of their employees and special interests than to their clients.

Perhaps for analogous reasons, the first large-scale study showed similar inefficiency of large districts and large schools in 38 states.[27] The study showed no effect of per-student spending but significant effects of each state's average district and school size. Why? Consider Montana: Usually at the top of state achievement surveys, its many districts have as few as 100 to 200 students, so school board members may be able to speak insightfully about many of the individual faculty and students in their single school. In New York City, board members might be stumped to name more than 50 of the roughly 900 schools of the 1.1 million students. If something goes wrong in a Montana school, a parent might ask a school board member at a grocery store to look into it. Can this be imagined New York City?

Teachers in the tiny Montana district, to continue the example, would be likely to know not only the students but also their siblings and other relatives. Parents, teachers, and school board members can readily communicate. Being small, neither the district nor the school would multiply programs and courses excessively, but they would stick to fundamental subjects in a core curriculum taken by most students, such as English, mathematics and science, civics, history, and geography, foreign language, and art and music, which has been shown to be conducive to high achievement and advantaged university admission. In the 1990s, several dozen statistically controlled studies showed the achievement advantages of small schools, which tend to be concentrated in small districts.[28]

Choice

Though controversial, the weight of the evidence favors achievement benefits of choice. U.S. experiments appear to show that vouchers have thus far only benefited

---

[27] Herbert J. Walberg and Herbert J. Walberg III, "Losing Local Control," *Educational Researcher* 23, no. 5(1994): 23-29.

[28] See a comprehensive review of many later school size studies by Valerie E. Lee, Anthony Bryk, and J. B. Smith, "The Organization of Effective Secondary Schools," ed. Linda Darling-Hammond *Review of Research in Education* 19 (1993), 171–268. Fewer studies have been made of district size.

African American students. Reviews by economists of over 35 studies show that the greater the degree of private choice in geographical areas, the greater the achievement and the lower the costs of all students. Finally, perhaps the largest and most sophisticated study ever undertaken shoed that the greater the degree of partental choice in nations, the greater the achievement levels.[29] Thus, indicators of state choice policies as well as the kind of school a NAEP respondent attends might be valuable given the recent Supreme Court decision, keen policy interest, and current controversies about it.

Tracking and Grade Retention

Grouping students reflects common sense. If students with similar levels of knowledge and skills are grouped together, teachers can avoid teaching them what they already know and what they are yet incapable of learning; with instruction more suited to them, students should find learning more efficient and pleasant. What forms can such grouping take and what are the achievement and other effects?

Developing Prerequisites

A rigorous long-term study suggests that children at risk of school failure because of poverty appear to benefit from high quality, academically focused preschools that prepare them learning in kindergarten and subsequent grades. Closer in preparation to middle-class children, such better-prepared children may continue to benefit as late as early adulthood. Many other studies, however, show no effects or quick fade out of early gains. In any case, the best designed and conducted study supports the general principles of early preparedness but may also suggest that children made more homogeneous in achievement preparedness learn faster.

Grade Retention

By itself, retaining students in grade appears ineffective. On the other hand, "socially promoting" unqualified students may give them and their classmates little reason to study. This policy, common in big cities, probably devalues the high-school diplomas of qualified graduates in the eyes of employers and others.

Chicago's Summer Bridge program for failing children threatened grade retention and provided intensive academic summer school. Though some students failed, the program showed outstanding effects; it was not only effective but also highly cost-effective. Thus, preschool and summer bridge programs tend to homogenize student achievement, that is, bring laggards up to others' achievement levels, which probably contributes to more effective learning for both groups.

---

[29] Dan Goldhaber, "The Interface between Public and Private Schooling," *Improving Educational Productivity*, ed. David H. Monk, Herbert J. Walberg, and Margaret C. Wang (Greenwich, CT: Information Age, 2001), 47–76; Clyde R. Belfield and Henry M. Levin, *The Effects of Competition on Educational Outcomes: A Review of US Evidence* (New York: National Center for the Study of Privatization in Education, Teachers College, Columbia University, 2001); Ludger Woessmann, "Why Students in Some Countries Do Better," *Education Matters*, Summer 2001, 65–74.

Classroom Grouping

Widely used in elementary schools, homogeneous achievement grouping within classes has small, positive effects (about .25 on average). In the later elementary grades, Matthew effects (of the rich getting richer) have typically caused wide variations in student achievement; a sixth grade may have third- and ninth-grade readers. Probably for this reason, the "Joplin plan" of bringing like-ability students from different classes and grade levels into homogeneous groups has larger effects (about .35) than within-class grouping. As identified by ability or achievement tests, highly able students benefit from "enrichment" of their studies, that is, the provision of greater depth of regular grade-level content (.40). "Accelerated" homogeneous high-ability classes that allow students to study advanced-grade material benefit them greatly (.90).[30]

Tracking

By high school, student achievement levels differ more widely, and most American high schools practice tracking; about 86 percent of high schools, for example, track mathematics classes. Some scholars urge "de-tracking,"[31] that is, heterogeneously grouping all high-school classes, but surveys "show solid support for tracking among parents, teachers, and students" Research on de-tracking is insufficiently rigorous to support the policy,[32] a good reason for NAEP items on these and the other proven, possible, and controversial practices and policies described here.

Class Size and Student/Teacher Ratio

Student–teacher ratios fell from about 27 in 1955 to 17 in 1997,[33] which accounts for much of the substantial rise in per-student expenditures. More teachers, however, do not necessarily make for smaller classes, since they may perform administrative and special duties, especially in large cities with substantial federal programs, which may require much bureaucracy. In any case, as pointed out in the opening section, student achievement remained stagnant, despite the sizable investment in more teachers per student.

The first meta-analysis of education research on class-size effects on achievement suggested a small beneficial effect of class-size reductions. The biggest apparent effects were in reductions below class sizes of ten; classes between 15 and 35 students differed very little in achievement. Few studies had been made of classes between 8 and 15,

---

[30] James A. Kulik, "Tracking, De-Tracking, and Skill Grouping," in *Can Unlike Students Learn Together?* eds. Arthur Reynolds, Margaret C. Wang, and Herbert J. Walberg (book submitted for publication). See also other related chapters.

[31] Jeannie Oakes and Martin Lipton, "Can Unlike Children Learn Together?" in *Can Unlike Students Learn Together?* ed. Arthur Reynolds, Margaret C. Wang, and Herbert J. Walberg (book submitted for publication).

[32] Thomas Loveless, *The Tracking and Ability Grouping Controversy* (Washington, DC: The Fordham Foundation, 1998).

[33] "Table 64," in *Digest of Education Statistics* (Washington, DC: U.S. Department of Education, 1997).

because classes in this range were rare and prohibitive in cost. In any case, the overall effect of class-size reduction appeared to be much smaller than the use of effective teaching methods.[34] Even a small effect was disputed. Large-scale studies, mostly by economists, showed no consistent effect of class-size reductions.[35]

A much-noted Tennessee experiment seemed to show an effect of reduced class sizes,[36] even though a single study may not outweigh the inconsistent results of many other studies. Even at face value, moreover, the Tennessee study showed a very small effect, limited to kindergarten. Continuing exposure to smaller classes in first through third grade showed no advantage, and returning students to normal-sized classes in fourth through sixth grades showed no harm. So reduced class size apparently only benefited kindergartners, and changes in class sizes did not affect achievement in the six later grades.[37]

In addition, the apparent effect was not of class-size reduction alone but accompanying monetary incentives for increased student achievement. Thus, the apparent small transient effect may be attributable to smaller class sizes, monetary incentives, or a combination of these factors.

A more recent large-scale natural experiment on all Connecticut elementary schools overcomes limitations of the Tennessee research. It is perhaps the most comprehensive study ever made of the class-size question, because it measured the effects of natural changes in class sizes from 10 to 30 students over two decades. It showed no class-size reduction overall, nor any at the upper or lower range of class-size reduction, nor in the earlier or later grades, nor for disadvantaged or middle class students.[38]

---

[34] Gene V Glass and Mary Lee Smith, "Meta-Analysis of Research on Class Size and Achievement," *Educational Evaluation and Policy Analysis* 1, no. 1 (1979): 2–16.

[35] Eric A. Hanushek, "The Evidence on Class Size," in *Earning and Learning: How Schools Matter*, ed. Susan E. Mayer and Paul Peterson (Washington, DC: The Brookings Institution, 1999), 131–168; Julian Betts, "Is There a Link between School Inputs and Earnings?" in *Does Money Matter?: The Link between Schools, Student Achievement, and Adult Success*, ed. Gary Burtless (Washington, DC: The Brookings Institution, 1995). In addition, Asian classes, which have as many as 60 students, usually rank at the top of international achievement surveys.

[36] Jeremy D. Finn and Charles M. Achilles, "Answers and Questions about Class Size: A Statewide Experiment," *American Educational Research Journal* 27, no. 3 (1990): 557–77.

[37] Eric A. Hanushek, "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis* 21 (1999): 143–164.

[38] Caroline M. Hoxby, *The Effects of Class Size and Composition on Student Achievement* (Cambridge, MA: National of Bureau Economic Research, 1998). Also available at choxby@harvard.edu.

What would happen if a state concentrated resources on reducing class sizes? California policymakers did just this at a cost of about $5 billion per year from 1996 through 2001. About two-thirds of California school districts took money from libraries, art, music, and maintenance to reduce class sizes in the first three grades. After three years, evaluators could infer no achievement effect of class size reduction. As they concluded, "There is no clear relationship between changes in the amount of exposure to CSR [class-size reduction] and changes in the average level of achievement. Increased exposure is not associated with greater gains in achievement." [39]

In view of definitively inconsistent research and California's experience, further class-size reductions seem unpromising. Such reductions, moreover, have been exceedingly costly. They are even more costly today, since student/teacher ratios have already been cut massively in recent decades. Reducing class size, for example, by a single student from 15 to 14 incurs more than twice the teaching costs of a single-student reduction from 35 to 34, even aside from the costs of new classrooms. In any case, since class size reductions are so costly and their effects so uncertain, NAEP might usefully collect data on them.

Widespread, Unsubstantiated Programs

School board members and most educators lack education and experience in accountability, and evaluation, and methods of psychometrics and statistics that would enable them to choose effective, efficient programs and weed out others. Though these tasks should be central to leaders aiming to measure, evaluate, and improve learning, they are neglected. Consequently, popular programs are often chosen by fad and reputation rather than by a careful review of evidence of their results and costs. Developers themselves often evaluate their own programs, but they may have neither the professional skills nor the disinterestedness to evaluation them properly.

Independent evaluations, for example, support the effectiveness of Accelerated Reader but show little effect of the popular Success for All. Reading Recovery has shown positive effects but at huge expense. There is little evidentiary basis for New American Schools programs, the Annenberg Project, and those Congress advocated in the Obey-Porter legislation. Yet, Congress, school boards, foundations, and firms have contributed hundreds of millions of dollars to support their implementation and operation.

The Title I Program for Students in Poverty

The federal government spent about $125 billion on Title I and now allocates about $8 billion annually. The program was to have reduced the gap between middle-class

---

[39] Brian M. Stecher and George W. Bohrnstedt, *Class Size Reduction in California: Findings from 1999–00 and 2000–01* (California Department of Education, 2001). Quote on p. 90. In an interview, the first author, a senior social scientist at RAND Corporation, said, "It would be nice if we could give an unequivocal answer to the achievement question. Then people could decide if the benefits were worth the costs. Unfortunately we can't." (RAND Corporation press release, February 4, 2002). The authors have yet to reach a final conclusion and will continue to study the initiative, but their latest finding corroborates the pervasive elusiveness of the class size effect.

students, often Anglos in suburbs on the one hand, and on the other, poor students, often African Americans and Hispanics in cities and rural areas. Congressionally mandated and independent studies show that the Title 1 program, even after three decades, did not diminish, much less eliminate, the poverty gap. A synoptic evaluation of Title I points to the lack of evidenced-based policies, practices, and programs. [40]

Teacher Selection and Compensation

Public-school teachers' salaries have long been chiefly determined by whether they are certified, their years of teaching experience, and their degree level, commonly a bachelor's or master's. Despite thousands of doctoral dissertations in education written each year, little solid evidence shows these salary determinants promote student learning. In fact, studies by labor economists suggest that verbal ability, knowledge of the subject matter, and graduation from a selective college are at least as important as the usual salary determinants.

To investigate the contribution of the contending measurable teacher attributes to student learning, the following equation could be estimated:

> Student achievement = Student input + teacher experience + teacher education + teacher verbal ability + teacher pedagogical knowledge + teacher subject-matter knowledge + teacher certification

In this equation, student input is indexed by previous achievement and demographic characteristics such as poverty, verbal ability is indexed by verbal tests or college selectivity or reputation, knowledge is measured by tests or course completion in the subject matter such as science, and a weight is estimated for each factor.

No study, however, comes close to this equation. The consequence of flawed studies is misleading implications for teacher certification, hiring, retention, and compensation. For example, simply showing that the students whose teachers have a master's degree achieve better may reflect not the learning advantage of a master's but the fact that teachers who are more experienced are more likely to have master's degrees and vice versa. Similarly, failing to take previous student achievement and demographics into consideration may mean that an apparent connection between experience and achievement is attributable to teachers transferring to middle-class schools that achieve well in any case. Estimating the equation above would test these and other causal possibilities.

A limited standard of proof calls for including prior achievement and student demographics in testing the possible influences of the other factors one or two at a time. A recent search uncovered only 18 such studies, nearly all by economists. These studies

---

[40] George Farkas and L. Shane Hall, "Can Title I Attain Its Goal?" in *Brookings Papers on Education Policy*, ed. Diane Ravitch (Washington, DC: Brookings Institution, 2000), pp. 59–123. Other problems with Title I include (1) measuring poverty, (2) the possible conflict of interest between educators who seek additional funds and families and students who may not wish to be identified as poor, and (3) concentrating Title I services on only poor children while trying to avoid the administrative problems and possible stereotyping harms of segregating them.

suggested that college selectivity, verbal test scores, and, only for high-school students in mathematics, subject-matter knowledge contribute to student learning.[41]

Examining studies that control only for student input trades a larger pool of studies for research rigor. Since teacher effects and costs are so critical, even less certain evidence is worth considering. Such research corroborates the importance of verbal facility and college selectivity, but suggests that only 3 percent of teachers' contribution to student learning is attributable to teacher experience and graduate degree attained. Few studies show positive, significant effects of experience and education, and some studies show significant, negative effects. Certified teachers apparently perform no better those who are uncertified.[42]

Even though teaching comprises about half the total schooling costs, research provides no support for traditional and current policies of certifying, selecting, and compensating teachers. Arbitrarily excluding candidates on weakly predictive or nonpredictive criteria is arbitrary; in an apparently tight labor market, this longstanding policy unduly excludes large numbers younger and older people who are as likely to teach as well as other candidates and the present labor force. In fact, Teach for America demonstrates that very recent graduates of elite colleges, knowledgeable of the subjects they teach, but with no experience and little pedagogical training, are highly regarded by their principals and that they also induce greater achievement than other teachers.[43]

Selecting NAEP Informational Items

Obviously, NAGB cannot include all items that bear upon the many causal factors identified and explained here if tradition is followed. But NAGB faces a rare opportunity to help the nation in the substantial achievement and accountability gains called for in the No Child Left Behind Act. It may be time to think again about how NAGB can add the most value to the nation's efforts to measure, increase, and monitor achievement and better understand its causes. Though NAGB by law must make the difficult decisions, what old and new principles seem reasonable, even though some trade-off against one another and may require additional (or less) expense and effort?

1. Minimize the number of informational items to avoid "response burden" (time and energy to respond to NAEP items), additional NAEP administering, processing, analyzing, and reporting, and to avoid jeopardizing NAEP's primary

---

[41] Andrew J. Wayne and Peter Youngs, "Teacher Characteristics and Student Achievement Gains: A Review" (paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Washington, DC, 2001).

[42] Dan Goldhaber, "The Mystery of Good Teaching," *Education Next*, Spring 2002, 50–55.

[43] Wendy Kopp, "Ten Years of Teach for America" in *Tomorrow's Teachers*, ed. Margaret C. Wang and Herbert J. Walberg (Richmond, CA: McCutchan, 2001), 221–234; Margaret Raymond and Stephen Fletcher, "Teach for America: The First Evidence on Classroom Performance," *Education Next*, Summer 2001, 62–70.

mission of measuring achievement because of controversies that might arise about the relevance and appropriateness of the items.

2. Retain old NAEP items and incorporate those from other major international and national surveys such as HSB, TIMSS, NLS, and NELS to enable useful comparisons with the past and with other countries. This would allow scholars, for example, to ascertain whether new federal or state legislation leads students do more homework or have their essays closely reviewed and annotated, known causes of achievement. In other words, NAEP would not prove causality but would allow policy makers to monitor changes in not only achievement but, as established by other research, its causes.

3. "Rotate" items among students. That is, just as NAEP now administers "testlet" booklets to groups of students within a class, any given student may be given only a small set of the items, although a limited number of required reporting items might be given to all students. (Numerically small samples of, say, a class or school are less at issue when they represent a substantial fraction of the population.)

4. Teachers might also be given rotated questionnaire items, which save their time, attention, and energies.

5. During the NAGB standard setting process, professors, teachers, and others who identify content for testing might also identify effective practices in their subject such as laboratories in science to identify important correlates and possible causes. In a science assessment, moreover, only the science teachers could be asked to respond to items in their subject.

6. State department, central-office, and non-teaching school staff might provide information about their bailiwicks.

7. Items might be chosen that bear on causal disputes so that researchers can try to shed light upon their effects.

8. Items might be chosen that have large, consistent, less disputed effects so that policy makers can compare states and trace their implementation and researchers can use them to take them into account when estimating less creditable factors just as an epidemiologist might control for age, occupation, and smoking in a population estimating the effects of pollution on emphysema.

Table 1

Nine Educational Productivity Factors

Factor	Proportional Learning Influence
A. Student Aptitude	
1. Prior achievement	.92
2. Development as indexed by chronological age or stage of maturation	.51
3. Motivation or self-concept as indicated by personality tests or the student's willingness to persevere intensively on learning tasks	.18
B. Instruction	
4. Amount of time students engage in learning	.47
5. Quality of the instructional experience, including method (psychological) and curricular (content) aspects	.18
C. Psychological Environments	
6. Morale or student perception of classroom social group	.47
7. Home environment or "curriculum of the home"	.36
8. Peer group outside school	.20
9. Minimal leisure-time mass media exposure, particularly television	.20

Source: Barry J. Fraser, Herbert J. Walberg, Wayne W. Welch, and John A. Hattie, "Synthesis of Educational Productivity Research," *International Journal of Educational Research* 11 (1987): whole issue. The estimates are calculated from data reported on p. 220. The indexes in the table are on the same scale as the effect sizes reported later but are not necessarily pure, one-way causal effects.

Table 2

Instructional Effects

Category	Average Effect
1. Identifying similarities and differences	1.61
2. Summarizing and note taking	1.00
3. Reinforcing effort and providing recognition	.80
4. Homework and practice	.77
5. Nonlinguistic representations (e.g., maps and other graphics)	.75
6. Cooperative learning	.73
7. Setting goals and providing feedback	.61
8. Generating and testing hypotheses	.61
9. Activating prior knowledge	.59

Source: Robert J. Marzano, *A New Era of School Reform: Going Where Research Takes Us* (Aurora, CO: Mid-Continent Research for Education and Learning, 2000), 63.

Note: The effects in this and other tables are generally ordered from largest to smallest as indicated by the effect sizes.

Table 3

Selected Effects of Quality of Instruction

General Methods	
Instructional Elements	
Cues	1.25
Reinforcement	1.17
Corrective Feedback	.94
Engagement	.88
Mastery Learning	.73
Computer-Assisted Instruction	
For Early Elementary Students	1.05
For Handicapped Students	.66
Teaching	
Comprehension Teaching	.71
Direct Teaching	.55
Teaching Techniques	
Homework with Teacher Comments	.83
Graded Homework	.78
Frequent Testing	.49
Pretests	.48
Adjunct Questions	.40
Goal Setting	.40
Assigned Homework	.28
Explanatory Graphics	.75

Special Methods	
Reading Teaching	
Adaptive Speed Training	.95
Phonemic Awareness	.86
Repeated Oral Reading	.48
Phonics	.44
Writing Teaching	
Inquiry	.57
Scales	.36
Sentence Combining	.35
Early Education Programs	
Preschool	.22–.50
Full-Day vs. Half-Day Kindergarten	.48
Grouping	
Acceleration of Gifted Students	.88
Tutoring	.40
Staff Development	
Feedback	.70
Staff Development for Reading Teaching	.61
Microteaching	.55

Source: Herbert J. Walberg and Jin-Shei Lai, "Meta-Analytic Effects for Policy" *in Handbook of Educational Policy*, ed. Gregory J. Cizek (San Diego, CA: Academic Press, 1999

Table 4

Student and Family Influences

Variable	Effect
Prior Knowledge	1.43
Motivation	.73
Family Background	
Home Environment	1.42
Parental Income	.67
Parental Occupation	.42
Parental Education	.38

Source: Robert J. Marzano, *A New Era of School Reform: Going Where Research Takes Us* (Aurora, CO: Mid-Continent Research for Education and Learning, 2000) 69, 70, and 73.

Note: The indexes in the table are on the same scale as the effect sizes in Table 3 but are not necessarily pure, one-way causal effects.

Table 5

School-Level Possible Influences

Variable	Effect
Opportunity to Learn	.88
Time	.39
Monitoring	.30
Pressure to Achieve	.27
Parental Involvement	.26
School Climate	.22
Leadership	.10
Cooperation	.06

Source: Robert J. Marzano, *A New Era of School Reform: Going Where Research Takes Us* (Aurora, CO: Mid-Continent Research for Education and Learning, 2000), 56.

Note: The indexes in the table are measured on the same scale as the effect sizes in Tables 3 but are not necessarily pure, one-way causal effects.

Table 6

Possible Policy Effects
Probably Worth Considering for Measurement

Policy	Effect
Curriculum based external examinations	++
Accountability	++
Choice	++?
Small schools	+
Small districts	+?
Small classes	+?
Tracking	0?
Grade retention	0?
Professional standards	0?
Title I	0?

Note: The number of pluses indicates the likely magnitude of the effect; a question mark indicates controversy and uncertainty.