

1 JOHN F. DAUM (SB #52313)  
2 FRAMROZE M. VIRJEE (SB #120401)  
3 DAVID L. HERRON (SB #158881)  
4 PAUL B. SALVATY (SB #171507)  
5 O'MELVENY & MYERS LLP  
6 Embarcadero Center West  
7 275 Battery Street  
8 San Francisco, California 94111-3305  
9 Telephone: 415.984.8700

10 Attorneys for Defendant State of California

11 **SUPERIOR COURT OF THE STATE OF CALIFORNIA**

12 **CITY AND COUNTY OF SAN FRANCISCO**

13 ELIEZER WILLIAMS, et al., ) Case No. 312 236  
14 )  
15 Plaintiffs, ) Date Action Filed: May 17, 2000  
16 )  
17 vs. )  
18 )  
19 STATE OF CALIFORNIA, DELAINE )  
20 EASTIN, State Superintendent )  
21 Of Public Instruction, STATE )  
22 DEPARTMENT OF EDUCATION, STATE )  
23 BOARD OF EDUCATION, )  
24 )  
25 Defendants. )

26 )  
27 )  
28 STATE OF CALIFORNIA )  
29 )  
30 Cross-Complainant, )  
31 )  
32 vs. )  
33 SAN FRANCISCO UNIFIED SCHOOL )  
34 DISTRICT, et al., )  
35 )  
36 Cross-Defendants. )

37 **EXPERT WITNESS DECLARATION RE ANITA A. SUMMERS, Ph.D.**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

I, Paul B. Salvaty, declare as follows:

1. I am an attorney with the law firm of O'Melveny & Myers LLP, counsel of record herein for defendant State of California ("the State").

2. The State has provided a list of persons whose expert opinion testimony the State intends to offer at trial of this action, either orally or by deposition testimony. The list includes Dr. Anita Summers, to whom this declaration refers.

3. Dr. Summers has agreed to testify at trial.

4. Dr. Summers will be sufficiently familiar with the pending action to submit to a meaningful oral deposition concerning the specific testimony, including any opinions and their bases, that Dr. Summers is expected to give at trial.

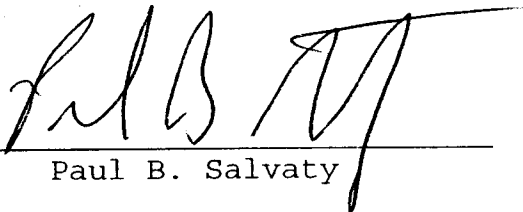
5. Dr. Summers' fee for providing deposition testimony, consulting with the State, conducting research and other activities undertaken in preparation of the attached report is \$400 per hour.

6. Pursuant to Section 2034(f)(2)(A) of the California Code of Civil Procedure, attached hereto as Exhibit A and incorporated herein by reference is a *curriculum vitae* providing Dr. Summers' professional qualifications.

1           7. Attached hereto as Exhibit B and incorporated herein by  
2 reference is Dr. Summers' expert report. Pursuant to Section  
3 2034(f)(2)(B) of the California Code of Civil Procedure, the  
4 following is a brief narrative statement of the general substance  
5 of the testimony that Dr. Summers is expected to give at trial.  
6 Dr. Summers rebuts the opinions of several of plaintiffs'  
7 experts, sets forth the ideal features of a state accountability  
8 program; evaluates the extent to which the systems in the 50  
9 states possess those features; and then concludes that, contrary  
10 to plaintiffs' assertions, California has one of the best  
11 accountability programs in the country. The foregoing statements  
12 are only a general summary of the issues and conclusions  
13 discussed and documented more fully in Dr. Summers expert report.

14  
15           I declare under penalty of perjury that the foregoing  
16 is true and correct.

17  
18           Executed this 18th day of April, 2003, at Los Angeles,  
19 California.

20  
21             
22           Paul B. Salvaty

## Anita A. Summers

Anita A. Summers is Professor Emeritus of Public Policy and Management in the Wharton School of the University of Pennsylvania. She has secondary appointments in the Graduate School of Education and the Real Estate Department. She is a Senior Research Fellow in the Wharton Real Estate Center, and member of the Brookings Institution project on Metropolitan Solutions to Urban and Regional Problems.

Professor Summers did her graduate work at the University of Chicago and Columbia University, and received an Honorary Degree from Hunter College in 1995. At the University of Pennsylvania, she was Chair of the Public Policy and Management from 1983-1988. She has been a Visiting Professor at the Hebrew University of Jerusalem, Temple University and the University of Hawaii, she was a Resident Scholar at the Rockefeller Foundation Study Center in Bellagio, Italy in 1986; and was Research Officer and Head, Urban Research Section of the Federal Reserve Bank of Philadelphia from 1975-1979. She was on the Provost's Academic Planning and Budget Committee from 1984-1990, Member of the Law School and Wharton School Dean Search Committees, and Chair of the Research Policy Committee at the University of Pennsylvania. She is currently Ombudsman for the University of Pennsylvania.

Professor Summers' research work has been in two major areas--education efficiency and urban economic development and finance. She has published extensively in both. Her most recent works, following the publication of three books on the Philadelphia metropolitan area are Urban Change in the U.S. and Western Europe: Comparative Analysis and Policy (edited with Paul Cheshire and Lanfranco Senn, published by the Urban Institute Press in 1993, second edition in 1999), Schooling and Labor Market Performance (with Amy W. Johnson and David Crawford, Economics of Education Review, 1997), contributor to Making Schools Work (Eric A. Hanushek, et al, The Brookings Institution, 1994), and "Accountability Gains: Expert Measures", Education Next, 2002. She recently completed working on a project, funded by the Pew Charitable Trusts, on Regionalization: The Feasible Options, has co-authored the first Brookings Institution's Policy Briefs- Paying for the Poor: A New Strategy for Handling the Urban Burden, and is co-authoring a chapter in the forthcoming Brookings volume, Forging Metropolitan Solutions to Urban and Regional Problems.

Professor Summers was on the Board of Directors of The William Penn Foundation, and is currently Chair of the Board of Directors of Mathematica Policy Research, Inc. She was a member of the National Panel on the Economics of Educational Reform, the Tax Policy Advisory Committee for the City of Philadelphia, and the Metropolis Project of the American Academy of Arts and Sciences; and is currently part of the Brookings Institution Urban and Metropolitan Policy Center Project. She has consulted with several local school districts on the design and implementation of assessment schools, and has been an expert witness on school finance for a number of states and school districts.

**ASSESSMENT IN EDUCATION:  
OUTPUTS AND INPUTS**

Expert Report of Anita A. Summers  
With the assistance of Gary W. Ritter

Prepared for Williams et al. v. State of California

April 8, 2003

## TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
I. Background	2
II. Introduction	4
III. Outputs	6
Useful Outputs for School Assessments	6
Objections Over Education Output Measures	8
Psychometric Problems	8
Teaching-to-the-Test	10
Calculation of Outputs	11
Use of Controls	11
Controls are a “Black Box”	13
Standards v. Absolute Measures	13
Summary re Output Measures	14
IV. Inputs	17
Role of Inputs in Educational Assessment	17
Which Inputs Matter?	18
V. Educational Outputs and Inputs: Usage in the 50 States	21
Table 1: Summary of Output and Input Usage in Educational Assessment in the 50 States	22
Table 2: Output and Input Usage in Assessment in the 50 States	25
VI. Conclusions: California’s Assessment Program	42
Footnotes	45
Bio: Anita A. Summers	46
Bio: Gary W. Ritter	47

# ASSESSMENT IN EDUCATION: OUTPUTS AND INPUTS

Anita A. Summers  
With the assistance of Gary W. Ritter

## I. BACKGROUND

The purposes of this report are two-fold:

- (1) First, the input and output characteristics of an ideal state educational assessment plan are described. The ideal characteristics are derived from the peer-refereed research literature of objective scholars in the field. The discussion is pointed to the issues raised in Williams et al. v. State of California, et al. by several of the plaintiff's expert reports.
- (2) Second, the current California system is placed in the national context. Details of the input and output characteristics of the educational assessment plans of the 50 states are arrayed in a large table. How California's plan ranks in comparison is discussed.

I am a Professor Emeritus in the Department of Business and Public Policy in the Wharton School of the University of Pennsylvania, with and have had a secondary appointment as Professor in the Graduate School of Education. I was responsible for what has been regarded as a seminal work on education production functions published in 1976 in the American Economic Review. Since then I have published a number of articles on educational efficiency, worked with several school districts on assessment plans, and have been an expert on several state and school district cases. I am currently Chairman of the Board of Directors of Mathematica Policy Research, Inc., one of the country's most respected evaluators of policy interventions.

The paper was written with the assistance of Professor Gary W. Ritter. He did all the research on assessment plans in the 50 states, and arrayed the input and output information in the

large table in Section V. Professor Ritter is an Assistant Professor of Education and Public Policy at the University of Arkansas. He has published extensively in peer-reviewed journals on subjects related to evaluation and assessment.

Fuller biographies are at the end of the paper.



## II. INTRODUCTION

The sharp focus on educational reform in the United States over the last quarter of a century has inevitably led to an intensified interest in assessment techniques and results. Without “the facts”, how can the nation, the state, the school district, or the parent know whether students are performing “well” in terms of their objectives? How can they know which factors account for high and low performances without the analyses? Education has joined many of the major public policies – criminal justice, job training, public health, and welfare – in its efforts to assess how effectively the system works.

A substantial portion of the assessment efforts is, however, devoted to reporting results in terms of outputs only. The plaintiffs in *Williams et al. v. State of California* argue for emphasis on the quantity and distribution of educational resources (inputs). Reform measures cannot (not just should not) be deduced from analyzing either inputs or outputs alone. Inputs are relevant only insofar as they affect desired outputs. They are needed as controls for calculating outputs, as a way of separating out factors controllable by the schools from factors affecting the output that are given to them. Outputs of individual components of the educational system – the student, the teacher, the principal, and the school district – can only be determined when the appropriate inputs are factored in. If the educational performance (output) of a student is very poor, how can that be attributed to the teacher (one input) if the real cause is a very dysfunctional home (another input)? If the appropriate inputs are not included in the analysis, the output of – in this case – the teacher, will be incorrectly assessed. Outputs are the translation of the objectives of an educational system into useful measures. Inputs are important in order to assign the outputs to

the appropriate educational player. And, the relevance of an input in the educational system can only be determined in relation to outputs.

Output measures are essential to assessment. Without output measures, there is no satisfactory way to determine whether or not an input is relatively important in the educational process. Without input measures, there is no accurate way of measuring outputs. They are both required for informative assessments, but it is outputs that reflect whether we are achieving our objectives. School and non-school inputs are the factors controlling those outputs.

This paper will expand on the significance of outputs and inputs, and how they are combined in educational assessments. Section III discusses outputs – which measures are useful, objections to various measures, how they should be calculated. Section IV presents a discussion of inputs – why they matter, what the literature has to say on which inputs matter, and how they should be used. Section V presents a summary of what is happening in each of the states in each of their assessment programs – which inputs, which outputs, and which accountability consequences are being used. In Section VI the conclusions emerging from this national context and from the consensus of those involved in rigorous research on inputs and outputs are stated. In brief, California's assessment system ranks high in the nation on the merits of the plan it is currently using. There are improvements that would be beneficial, but the fundamentals of its usage of outputs and inputs puts it in the top group of state educational assessment systems.

### III. OUTPUTS

#### Useful Outputs for School Assessments

There is widespread agreement that some set of measures reflecting student cognitive achievement should be a dominant objective.<sup>1</sup> Almost all serious scholars agree that student achievement should be a dominant goal of any school system, a goal that should pervade all the processes of the system. They are generally agreed that articulating student achievement as a dominant goal provides a motivating focus for students, staff and administrators – and for teachers who are energized by this focus, because student learning is what they regard their job to be. Many argue for multiple measures of student learning, but, even then, it is understood that cognitive achievement is a necessary – if not sufficient – desirable result.

There are two other output measures that are commonly recognized as important measures of the educational output of a school. The dropout rate of a school is informative about the proportion of students who have not been caught up with the importance and life-time implications of learning. The proportion of students going on to post-secondary education is another important measure of the degree to which students were swept up in the learning process. It indicates the extent to which they perceive education to be a desirable way of investing in the development of their human capital.

All three of these measures – cognitive achievement, dropout rates, and post-secondary education – are important measures of student learning. An ideal assessment system will keep track of all three output measures for every school system, and regard them – together – as the basis for judging the quality of the learning process. It would also track the very significant indicator of the human capital of graduates of the education system – post-secondary school

earnings. In reality, however, although most assessment material discusses the use of these multiple measures, most states end up just using up test scores for several reasons: (1) The data are readily and regularly available, and there is implicit recognition of the importance of cognitive performance. (2) The basic skills are a necessary set of tools for the acquisition of knowledge and post-secondary performance. There is evidence that mathematics has become an increasingly essential tool in our technological society. (3) There is evidence that mathematics has become an increasingly essential tool in our technological society. (4) The results of cognitive tests have been shown to be significantly associated with their productivity and earnings.<sup>2</sup> (5) The collection of data on post-secondary performance involves an effort that is outside the traditional scope of school district data compilations. Students have to be tracked after they leave the system.

The literature of assessment is filled with many, many suggestions for measuring educational quality – good citizenship, high ethics, environmentally conscious, sense of community, appreciation of diversity, etc. These may all be desirable objectives for one's child, but several problems arise in using them as school outputs: Is the school the place that should be held responsible for developing these qualities? Do we have any satisfactory way of measuring the degree to which they are attained? How many things can we expect a school system to focus on – shouldn't we focus on the skills that are uniquely suitable to institutions of schooling?

There is a second major group of "outputs" suggested – student attendance, teacher attendance, availability of modern technology. Here the problem is that these are inputs, not outputs. If students do not attend, then their outputs will be low – cognitive skills will be poor, dropout rates will be high, and post-secondary education will be difficult to attain. Since teachers are documentably the most significant input in the school system, their poor attendance will

affect outputs – but it is not an output measure. Computer availability is a measurable input. The jury is out on the effect on cognitive learning, though a number of serious studies are underway. The evidence thus far suggests that technology appears to have a very weak effect on student achievement, but, in any event, it is not an output measure.<sup>3</sup>

### Objections Over Educational Output Measures

Though there is widespread agreement on the use of measures of cognitive achievement, dropout rates, and post-secondary schooling, when specific definitions of these are formulated for specific state assessment plans, considerable dispute usually arises. Why? There are several areas of contention. Some arise from political tensions, some from a simple lack of understanding of the statistical methods underlying the calculations.

Assessment activity has spread across the country and the evidence is unambiguous that test scores are always the prime focus (as is clear from the review of assessment activity in the 50 states in Section V below). The decibel level of objections to the use of test data has increased. Several objections are raised. There are responses to each of them:<sup>4</sup>

Psychometric Problems. There is concern over the imperfect capabilities of tests in assessing knowledge, the statistical noise in any scale used (each of which has defects), the compounding of the “noise” problem when two scores are subtracted to obtain a value-added measure, the difficulty of comparing the data of small and large schools, and the difficulty of comparing value-added assessments across subjects.

First, sophisticated users of assessment measures should not implement actions (using a new reading program, state government takeover of a school, rewards to outstanding schools) based on small numerical differences. No significant policy intervention should be implemented on the basis of small measurement differences – and they aren’t! Elementary school educators

know that if a student is reading below grade by the end of the fourth grade, that student will have considerable academic difficulty in many subjects in the subsequent grades. How do we know that? We know and act on that from repeated carefully done studies, not by one finely calibrated result. Texas, for example uses sophisticated econometric techniques involving several output measures to reward schools. They award a group of top-ranking schools, where the ranks are determined by the statistical analysis. They do not give the awards based on decimal point differences among schools on one test. This appears to be true for all the assessment schemes using sanctions and rewards. The “noise” should preclude decisions based on small differences. It should not preclude decisions based on gross findings, and states appear to understand this. (California’s use of decile bands for categorizing API results, with graduated consequences, reflects this recognition.)

Second, confidence in gross findings can be developed by replication, and by examining results over several time periods. Many important areas of educational research are replete with statistically sophisticated replications. Examples are the effects of class size, the education credentials of teachers, and instructional expenditures on student learning. Examining the stability of findings over several time periods is also a tool. If, for example, a particular school operates with great dedication and zeal, the finding that the students in that school have much higher performance than one would predict should not be a one-year finding. It should show its robustness by repeatedly coming out as a high performer. Now that assessment schemes are developing and getting used across the country, the replication results will be evident automatically.

Third, every effort should be made to reduce total dependence on test scores as an output measure. Other measures mentioned above – dropout rates, post-secondary education, and post-

secondary earnings – should be introduced into the assessment plan. The richer the measures used, the less weight there is on the psychometric concerns about test scores. But, if resources allow for only one measure, the undeniable necessity for learning the basic skills should lead to the selection of test scores. Several states do use dropout rates as an output measure, but only two use any surrogate for postsecondary school performance. (See Table 1 and 2 in Section V.)

Fourth, concerns have been expressed over difficulties involved in comparing the results of large and small schools, and comparing results of one subject with another. The difficulty only arises if the comparison groups are not well formulated. The most ideal unit of observation is the student. Innumerable studies have shown why the data should be collected at the student level:<sup>5</sup>

- (1) The same cohort of students can be followed without measurement errors arising from socioeconomic background and prior school effects.
- (2) The results can be aggregated into any units the assessors are interested in analyzing – the school, all fourth grade teachers, the classroom, all history teachers.
- (3) The confounding effects of student mobility can be addressed.
- (4) Comparisons can be made within one subject category – all high school science classrooms, for example.

Teaching-to-the-Test. This objection has been raised regularly. The argument is that, if the results of test scores are major assessment measures with consequences, teachers will alter their curriculum and teaching methods to maximize the test scores of their students. There are several responses to these concerns:

First, the function of tests is to communicate what is important to learn (established by the curriculum framework) — for both students and teachers. If tests are appropriately aligned with what states and school districts want their students to know, then the results of the tests will reflect the extent to which students have absorbed what their society wants them to learn. If that

is not the case, the problem is not the use of cognitive tests as an assessment tool. It is the use of an improper test.

Second, as discussed above, output measures for educational assessment should ideally use multiple measures – not just test scores. The ultimate question output measures are trying to assess is – what contribution is schooling, and specific segments of the schooling process (the teacher, the class size, the reading program, the computer lab) making to the development of the human capital of their students? The current emphasis on test scores as the prime output measure should translate into teachers and principals working to have students develop cognitive skills. And, indeed, this appears to be the case. There is clear evidence of improvement in basic skills in states with high stakes accountability systems.<sup>6</sup> There is every reason to expect similar results if, and when, additional outcome measures are emphasized.

Third, most of the research evidence indicates that low-achieving students are the largest beneficiaries of the installation of assessment programs. When they receive teaching-to-the-test exposure to the same body of knowledge over and over again, training on how to take a test under a time constraint – they benefit.

### Calculation of Outputs

Use of Controls. Responsible users of assessment results would not use a simple output measure as a tool for sanctions and awards. For example, in its simplest form, one might observe that the fourth grade math scores in School A are significantly higher than in School B. Does that mean that the “output” of School A is higher than School B – does it mean that the efforts of the teachers, principal and staff of School A was greater than that of School B – that School A was a “better” school? It might or it might not. It might mean that the students in School A



came from homes that had parents with more education and income than those in School B. It might have nothing to do with the school's efforts. The effects of all that the school gives the student can only be isolated by controlling for the socioeconomic status (SES) of the students. It might mean that the students in School A were better prepared in previous grades than the students in School B. The prior achievement levels should also be introduced as a control—and, indeed, almost all assessment schemes do this. It might mean that School A had access to a computer technology for teaching fourth grade math that School B did not have. In an ideal system, these other possible relevant inputs would be put in the calculation in order to isolate the effects of teachers and schools.

Most assessments include some sort of control for SES, frequently by presenting the test score results by SES categories. Many use a “value-added” methodology that controls for all the SES and prior factors that might determine a student's score in the current year by looking at the change in score from a baseline. Some use level changes, some use percentage changes.

The more complex forms of the output calculation – using student data and measuring changes in the performance of an identical cohort of students – would yield the most valid measures of school output.<sup>7</sup> The extent to which inputs such as class size, textbook availability, and teacher quality, explain the differences in test score changes could then be examined by introducing those variables.

Essentially, outputs cannot be measured satisfactorily without controlling for relevant inputs. Inputs are not “interesting” in and of themselves. They are important if they affect outputs. If there are school-controlled inputs that are identified as being significantly related to outputs, these inputs can then be identified as targets for improving student learning. Assessing the performance of school districts or schools by measuring the availability of selected inputs

alone is not productive in assessing student learning. If the input cannot be shown to be a statistically significant determinant of learning, what is it that is being assessed?

Controls are a “Black Box”. Many critics of value-added assessments argue that it is just too complicated for teachers, principals and the public to understand the results of using controls in the calculation of outputs. People who have not received some training in the use of statistical procedures to isolate the effects of an input on an output naturally enough find the concept difficult to understand. The problem is that, if transparency is the criterion that trumps all other criteria, we would be compelled to use assessment results that we know to be wrong. For example, there is no satisfactory way to make a judgment about which method of teaching is superior – whole language or phonics or a mixture – without factoring in the SES, and the school and teacher characteristics of each of the groups in the experiment. To find out which method is best, the results cannot be confounded with other factors that are affecting the final measures. We do not require a transparent understanding of the efficacy of a flu shot, or of the operating mechanisms of our car. We trust the experts. So must it be in educational assessment.

Standards v. Absolute Measures of Output. Output measures can be of two different types. They can be measured in absolute terms (a specific test score or change in test score over a period of time) or they can be measured in relation to a standard (absolute or percentage distance from a previously specified standard). There are clear problems associated with measurements in relation to a previously determined threshold. The major problem is one of setting the threshold. Those who want rewards for being at or over the standard opt for a lower output number. Those who focus on budget costs want to set a higher target. The threshold becomes a major bargaining tool, divorced from the objective of informed understanding of performance. A second problem relates to incentives. If a state is using an assessment system to create incentives for

improvement, as well as for informed decision-making, then there are distinct limitations in the usefulness of measures related to the attainment of standards. What incentive for improvement would there be for those who crossed the threshold? Absolute measurements provide the most complete set of information. Standards for rewards and sanctions can then be determined by the state or school district.

### Summary re Output Measure

Without output measures there is no assessment. The objective of educational assessments of school districts, schools and classrooms is to link the type and quantity of educational resources (the inputs) to the human capital results (the outputs). There are three major reasons for wanting this information: (1) to find out how our educational system is working – who is learning, who is not? how does it compare to the past? to other systems?; (2) to find out which inputs are most successful in producing the results the stakeholders in education want; (3) to provide incentives and assistance for improvement.

Specific definitions of outputs vary widely across the country, as is clear from Tables 1 and 2 in Section V below. The tables contain a summary of the assessment tools and methods used in each of the 50 states. The actual output measures in each state are products of the political process in state legislatures, the fiscal constraints of the state budgets, historical usage, the view of teacher unions, court precedents in school equity and finance lawsuits, and the educational research feeding into all of these. There are, however, certain characteristics of output measures on which there is substantial agreement among those doing independent research in this area. There is a consensus among them that an ideal educational assessment system would incorporate the following features:

- Regular, impartial measurement of the learning pattern in a class school, school district, and state are essential to improving educational quality – as they are essential in every major sector of our country’s socioeconomic activity.
- Cognitive scores on standardized tests are a necessary (though not sufficient) measure of student learning.
- Multiple output measures are to be preferred: cognitive scores, dropout rates, and some post-secondary performance measure are generally agreed upon indicators of educational quality.
- Data should be presented at the student level. Average data for a class or school or school district should not be an average of averages, but an average of the appropriate cohort of students.
- There is widespread agreement that all of the outputs should be measured in terms of change from one time period to another.
- Output measures, if they are to yield the most information, should be measured in absolute terms -- a specific test score, for example. If they are measure in relation to a standard only, much information is lost. A threshold standard is a blunt instrument, and the full information exists – unused, except to calculate the less precise measure!
- Outputs cannot be used satisfactorily without some level of controls. It is not possible to conclude that one school or school district is better at delivering education than another without knowing (factoring in) the socioeconomic levels of the students in each group and the previous educational accomplishments of each.
- Small differences in output measures should not be used to trigger sanctions or rewards. Random variations in performance and statistical errors of measurement

suggest that the output measures should be used in broad bands, as, indeed, many states do.

Most impartial assessment analysts, if they were not subject to political and interest group pressures, would agree on this list. The reality is, of course, that the assessment plans in use inevitably do reflect these pressures. As assessment protocols are increasingly developed and used, these basic “ideal” characteristics are, however, likely to prevail more widely.

## IV. INPUTS

A summary statement of the previous section is that outputs matter, inputs matter if they are significantly related to outputs, and there is much to be learned from sorting out which inputs are and are not related. There is no logic to a debate on whether assessments should be based primarily on outputs OR inputs. Meaningful information requires using both simultaneously.

### Role of Inputs in Educational Assessment.

Analysis of inputs is essential to the determination of the value added by the educational unit being assessed. If the assessment objective is to determine the relative contribution of different middle schools to the mathematics competence of its graduates, then the effects of other factors – the mathematical competence of the cohort of the middle school graduates when they entered the middle schools and the socioeconomic characteristics of those students are examples – must be controlled for. Without entering those control variables, the effectiveness of any particular school would be confounded with the effects of factors over which the school would have no control.

If the analysis yields the finding, for example, that, given the SES characteristics and prior mathematics achievement of the students, Schools A, B and C produced decidedly better results than Schools D, E, and F, then a further round of analysis remains to be done with the inputs. The next logical question is: Why do Schools A, B, and C produce better results? Which characteristics of their educational delivery accounts for their better output (mathematics scores) – textbook availability (if there are differences among the schools)? the age and condition of the middle school building, measured in some sort of uniform, independent way? the quality of the mathematics teachers? the use of programmed learning technology? In this round, the standard

procedure would be to establish whether or not there are differences in the quality of teachers that made a difference in the results by including T1, T2, T3..., one for each teacher. If, indeed, certain teachers show up as producing better mathematics scores in their classes than others, then, again, the question is why – more experience? more education credentials? more years of formal training in mathematics? No identifiable characteristic?

In sum, assessments require the analysis of inputs. Such analyses are needed in order to assess the impact of different units of educational delivery - schools, school districts, classes; they are needed to determine which inputs of the school system are relevant to the desired outputs of the school system; and they are needed in order to develop plans for improving or expanding the significant inputs.

### Which Inputs Matter?

The search for which school-controlled inputs matter to educational outcomes has been very, very extensive. The set of inputs not under the control of school systems – the family and peer group background of the student – has been clearly established as the input that has, by far, the strongest effect on student learning. Research in the field using sophisticated or simple methodologies, examining cognitive or noncognitive outputs, converges on the finding that the family background – parental education, income, and involvement in the education of their children – plays the most important role in the learning outcomes of their children. The strength of this result has affected the agenda for American schools over the last quarter of a century. Educational resources have been directed to providing items such as breakfasts and lunches, after-school child care, and remedial education to students in very low income families in an effort to offset the SES disadvantages so many students face. Unfortunately, the evidence

suggests that these initiatives have not been very successful in offsetting the handicaps of the family background. In educational policy, as in the areas of health and crime, the handicaps of impoverished and uneducated backgrounds have proven to be very intractable.

For school systems, however, the main interest centers around the effectiveness of the school inputs over which they have control. They do have control over the curriculum and associated materials, the school organization (more or less control at the school level, for example), the size of classes, the efficiency of the administrative operations, the maintenance of the physical plant, greater stakeholder involvement, the credentialing and other criteria in hiring teachers, and many other elements. The bottom line on what we know from conclusive studies about the effectiveness of these inputs can be summarized. Conclusive studies are defined as those based on large scale empirical analysis of hard data (not the opinions of students or teachers or parents or principals), studies that have been replicated in different places, and those that have stood the test of independent peer review. What we know<sup>8</sup> is that there is no conclusive evidence that the number of text books, the condition of the infrastructure of the school, greater stakeholder improvement, or the degree of school-based management has a significant impact on student learning. We know that the traditional credentialing characteristics of teachers that determine salary, the number of degrees and credits of education a teacher possesses, has no effect on the cognitive performance of their students; that smaller classes appear to have an effect on the learning of very young and very low achieving students; that having a few years of teaching experience has a positive effect; that having teachers with significant training in math and science has a positive effect; and that having teachers with degrees from higher-rated schools has a positive effect.



Most important, from study after study, we know that, though we cannot distinguish most of the characteristics that make for an excellent teacher (one whose students do better than predicted), we know that some teachers have significant positive effects and some have significant negative effects. The variation in what makes for an excellent teacher appears to be enormous, but we do know that an excellent teacher is the single most important school-controlled input, and that the credentialing characteristics that set salaries are not related to student performance.

The absence of conclusive evidence on many other inputs exists because full-scale empirical studies have not been mounted. The lack of impact infrastructure quality has been shown in one such study, but has not been replicated.<sup>9</sup> Evidence that fewer textbooks per pupil has a negative effect on educational performance cannot be deduced from a sample of teachers saying there are insufficient textbooks, or from a simple table showing that schools that have fewer textbooks per pupil have lower scores. Maybe the lower scores are due to the poor socioeconomic background of the students. Maybe they are due to students with lower scores losing more textbooks. Maybe there is a causal relationship. There is a way to find out. The right way is to do a serious empirical analysis, using textbook availability as one of the inputs, educational results as an output, and perform the appropriate statistical procedures. The wrong way is to buy more textbooks and then see if it makes a difference – or to deduce the answer from simple tabulations.

It is hard to understand why so many resources are expended on speculating about the role of various inputs, when there are established techniques for establishing the evidence.

## V. EDUCATIONAL OUPUTS AND INPUTS: USAGE IN THE 50 STATES

Educational assessment activity is taking place in some form or another in all 50 states. There are only three states — Montana, North Dakota, and Utah —that limit their activity to state testing requirements, but have no public exposure of the results. There are no designated units of assessment, outputs, inputs, or consequences.

California's assessment procedures need to be judged in terms of the national context, and in terms of the research standards discussed in the previous sections. California's system ranks high in these terms.

Table 1 below is a summary table of the extensive Table 2 at the end of this section. Information was accumulated on assessment activities in all 50 states. Published sources, web pages, and telephone calls were the source of all this information. Changes in state assessment activities can occur rapidly, and not all sources agree perfectly on the information. It is possible, therefore, that there may be some errors or changes in some of the details. In addition, the requirements of the No Child Left Behind legislation will undoubtedly affect the systems in many states.

In Table 2, the accumulated information for each state is presented. Col. (1) is the state name; Col. (2) gives the unit being assessed – the school, the school district, particular subgroups of students; Col. (3) identifies the output measure or measured used by each state; Col. (4) describes how the output measures are calculated; Col. (5) indicates whether or not any control factors were used to separate out school effects from nonschool effects; Col. (6) describes the specified consequences – sanctions or rewards – flowing from the assessment measures. Sources

**Table 1: Summary of Output and Input Usage in Educational Assessment in the 50 States**

<u>Unit of Aggregation in Output Measure</u>	<u>Number of States</u>	<u>Inputs</u>	<u>Number of States</u>
School <sup>1</sup>	47	No Rating System <sup>2</sup>	20
School District <sup>1</sup>	47	Rating System – No Inputs	5
Subgroup: Race	20	Prior Performance	24
Subgroup: Poverty	19	SES	4
Subgroup: LEP	21	School Characteristics	1
Subgroup: Special Ed	25		
Subgroup: Other	2		
		<u>Consequences</u>	
<u>Output</u>		None	3
No Rating System <sup>2</sup>	20	Sanctions Only	14
State has a Rating System	30	Rewards Only	1
Emphasis on Low-Perf.	14	Sanctions and Rewards	14
Standardized Tests	30	Neither Sanctions nor Rewards	21
Stand. Tests Only	15		
Stand. Tests and Other	15	Public Announcement	47
Dropout / Grad Rates	11	Financial Awards	15
Attendance	13	Improvement Assistance	27
Other	7	School Closure	11
		Reconstitution	19
		Student Transfers	11
<u>Output Measure</u>		Private Management <sup>4</sup>	6
No Rating System <sup>2</sup>	20	Withholding Funds	3
Level	29		
Change in Level	25		
Standard / Threshold <sup>3</sup>	13		
Continuous Measure <sup>3</sup>	21		

1. Missouri assessment procedures involving accreditation at the school district level only. Public report cards are required for individual schools.
2. No Rating System means that the state has test scores, but makes no formal attempt to rate the schools. Of the 21 with no rating system, 3 states do not require published school-level report cards; the other 18 do require school-level report cards but do not use the data in a rating system.
3. Standard / Threshold refers to dichotomous output measure in which a student either meets the proficiency standard or does not. The Continuous Measure refers to output measures that are based on continuous indicators or on categorical indicators with more than two categories.
4. State accountability system includes a policy for turning over the management of failing public schools to private/non-profit firms.

of the information are listed at the end of the table. Table 1 below is a summary of this full table. What are the most common characteristics of the state assessment programs currently with respect to their output measures, the input (control) measures, and the consequences flowing from the assessment results?

### Outputs

- Apart from the three states that, in effect, do not have any public assessment system, all the states aggregate their output measurements to the school and school district level. Somewhat less than half focus on one or more subgroups. Most commonly the categories are race, poverty, special education, and English as a second language.
- Some sort of rating system for all schools or school districts in the state exists in 30 states; 14 states only rate the performance of the lowest achieving schools; 20 of the states do not rank the schools or school districts at all. They make “Report Cards” public in 17 of these.
- Standardized tests are used by all 30 states that have a rating system for ranking schools and school districts—15 of them use them as the only output measure; 14 of them use multiple measures: Standardized tests and other measures - most often dropout and attendance rates. These are only two states that use some measures of postsecondary performance — Oklahoma and Kentucky.
- Of the 30 states with a rating system, 29 use an output measure of the level of performance (either a percentage or score measure) and 25 use a change in the levels. Many states use both. Thirteen states use a threshold or standard measure; 21 use a continuous (or continuous categories) as a measure; 4 use both.

### Inputs

- Most states with rating systems use some form of controls. They factor in inputs in order to appraise the outputs. Only 5 of the 30 do not use controls. Most common is the use of prior performance. When change measures (“value-added”) are used, the package of SES and previous school effects are wrapped in the prior performance measure and constitute a significant package of relevant inputs. Twenty-four states use that method.
- Only one state - California - uses some school characteristics as inputs (mobility, teacher credentials, class size, year-round educational programs).

## Consequences

- A credible assessment process should include some consequences associated with the findings. Only 3 states (the ones mentioned above) have no consequences.
- Of the remaining 47, all states use public announcement as a consequence, 15 use financial awards, 27 provide improvement assistance; as sanctions, 11 use school closure, 19 use reconstitution, 11 use student transfers, 6 use private management, and 3 withhold funds.
- In the economy at large, sanctions and rewards are used as incentives. Among the 47 state with assessment programs, 14 use sanctions alone, 1 uses rewards alone, and only 14 use both rewards and sanctions.

**Table 2: OUTPUT AND INPUT USAGE IN EDUCATIONAL ASSESSMENT IN THE 50 STATES**

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Alabama	School. School District. <u>Subgroups:</u> None (although online reporting system does present subgroup data)	Schools in immediate need of academic assistance have been identified as Academic Priority Schools based on state test scores.	Academic Priority Schools are selected based on three criteria: (1) Stanford Achievement Test average below the 30 <sup>th</sup> percentile. (2) No students met the academic content standards on Alabama Direct Writing Assessment. (3) Less than 80 <sup>th</sup> of 12 <sup>th</sup> grade students passed all parts of Alabama High School Graduation Exam.	None.	No Financial Rewards. Sanctions for Academic Priority Schools; sanctions include assistance and reconstitution. Publicly Available Report Cards.
Alaska	School. School District. <u>Subgroups:</u> None.	<i>No analysis; report card of test scores and other school data only.</i>	<i>No analysis; report card of test scores and other school data only.</i>		Publicly Available Report Cards.
Arizona	School School District <u>Subgroups:</u> None.	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Colorado	School. School District. <u>Subgroups:</u> None.	Overall Academic Performance is a single score (level) for each school based on state test scores. Academic Improvement is the growth in standardized weighted test scores from the previous year.	Overall Academic Performance is a single score (level) for each school based on weighted average on state test scores for several grade levels and subjects (based on five possible performance categories). The overall standardized weighted scores are then classified into one of five rating categories. Academic Improvement is the growth in overall standardized weighted test scores from the previous year, divided by the statewide standard deviation. Academic improvement scores are then classified into one of five rating categories.	Prior performance.	Rewards (financial) for high ratings in Overall Academic Performance or Academic Improvement. Sanctions for schools with low scores in Overall Academic Performance; sanctions include assistance, reconstitution, and student transfers. Publicly Available Report Cards.
Connecticut	School. School District. <u>Subgroups:</u> None.	Schools in immediate need of academic improvement, based on state test scores. (called Priority Schools).	Priority Schools are selected based on the fraction of students scoring within each of 3 performance categories on the state test.	None.	No Financial Rewards. Sanctions for Priority Schools; sanctions include assistance, closure, reconstitution, and student transfers. Publicly Available Report Cards.
Delaware	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	School Performance Rating is a single score (from 4 categories) for each school based on the level and growth on state test scores.	School Performance Rating is a single score (from 4 categories) for each school based on: (1) School Composite Score, based on absolute performance of students on state tests for several grades and several subjects. (2) Progress toward a state-mandated Target School Composite Score. (3) Progress in decreasing the % of students below standard.	Prior performance.	Rewards (financial) based upon Performance Status. Sanctions for schools labeled under student improvement; sanctions include assistance. Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
<b>Florida</b>	School. School District. <u>Subgroups:</u> (1) Lowest performing 25% of students in reading	School grade is a single score for each school based on both level and growth on state test scores.	School Grade (A through F) is a single score for each school based on the test scores for several grade levels and subjects. Grade is based on both level and growth: (1) Level – fraction of students scoring at proficient levels. (2) Growth – fraction of students making annual learning gains (with a particular focus on the lowest 25% of students in reading).	Prior performance.	Rewards (financial) for school grade of A or growth of at least one level. Sanctions for school grade of F; sanctions include assistance and student transfers. Publicly Available Report Cards.
<b>Georgia</b>	School. School District. <u>Subgroups:</u> (1) Special Education	<i>No analysis; report card of test scores and other school data only.</i>	<i>No analysis; report card of test scores and other school data only.</i>		Publicly Available Report Cards.
<b>Hawaii</b>	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	<i>No analysis; report card of test scores and other school data only.</i>	<i>No analysis; report card of test scores and other school data only.</i>		Publicly Available Report Cards.
<b>Idaho</b>	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	<i>No analysis; report card of test scores and other school data only.</i>	<i>No analysis; report card of test scores and other school data only.</i>		Publicly Available Report Cards.



(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Illinois	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.
Indiana	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.
Iowa	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Kansas	School. School District. <u>Subgroups:</u> (1) Special Education	Quality Performance Accreditation, based on state test scores and other measures of the school's choosing, is conducted for each school. (Accreditation is good for 5 years.)	The Quality Performance Accreditation process allows the school personnel to set their own goals, develop a school improvement plan, and satisfy the Kansas State Board of Education. The accreditation is based upon continuous improvement of all student groups in target subject areas on state tests and sustained effort measures (graduation rate, attendance rate, and other measures of the school's choosing).	Prior performance.	No Financial Rewards. Sanctions for schools denied accreditation; sanctions include assistance. Publicly Available Report Cards.
Kentucky	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	Academic Index is a single score (level) for each school based on test scores and 4 other factors. Accountability Trend (growth) for each school.	Accountability Index is a single score (level) for each school based on test scores on several grade levels and subjects (range from 0-140). Score is the weighted average based on the fraction of students scoring within each of four performance categories on state exams, in addition to attendance, retention, dropout rate, and successful transition to adult life. Accountability Trend is based on the amount of progress a school must make every two years to reach the State's 10-year goal score of 100. This is computed every other year (biennially).	Prior performance.	Rewards (financial) for meeting biennial trend goals that meet criteria regarding decreasing the number of lowest performers and dropout rate. Sanctions for schools below Accountability Trend line; sanctions include assistance and student transfers. Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Louisiana	School. School District. <u>Subgroups:</u> None.	School Performance Score (SPS) is a single score (level) for each school based on test scores, attendance, and dropout rate.  Progress of the Growth SPS toward Growth Target for each school.	School Performance Score (SPS) is a single score (level) for each school based on test scores on several grade levels and subjects (range from 0-100 and beyond). Score is the weighted average based on the fraction of students scoring within each of five performance categories on two tests, in addition to indicators of attendance and dropout rate.  Progress toward Growth Target is based on the amount of progress a school must make every two years to reach the State's 10- and 20-year goals.	Prior performance.	Rewards (financial) for meeting or surpassing Growth Target and showing growth in performance of high poverty and special education students.  Sanctions for low SPS and low growth; sanctions include assistance, reconstitution, student transfers, and withholding funds.  Publicly Available Report Cards.
Maine	School. School District. <u>Subgroups:</u> None.	<i>No analysis; report card of test scores and other school data only.</i>	<i>No analysis; report card of test scores and other school data only.</i>		Publicly Available Report Cards.
Maryland	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	School Performance Index (SPI) is a score (level) for each school based on state tests, attendance rates, and dropout rates.  Improvement is based on relative change between SPI 3 years prior and current SPI.	SPI is a score (level) for each school based on state tests, attendance rates, and dropout rates. The SPI is a weighted average of the school's relative distance from the satisfactory standards on state exams, attendance rates, and dropout rates. The SPI is calculated by dividing the school's performance on those indices by the state standards.  Improvement is determined by comparing the SPI for the current year against the SPI for the previous 2 years.	Prior performance.	Rewards (financial) for schools making growth.  Sanctions for schools not making adequate yearly progress; sanctions include assistance, closure, reconstitution, private management, and withholding funds.  Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Massachusetts	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	Performance Rating is a score (level) for each school (1-6) based on state tests (based on Proficiency Index, ranging from 0-100); there are two separate ratings for Math and English Language Arts.  Improvement Rating based on progress toward Improvement Target (5 categories). Ratings every 2 years.	Performance Rating is a score (level) for each school based on state tests (calculated separately for Math and English Language Arts). The Performance Rating category is based on the Proficiency Index (0-100), a weighted average based on the fraction of students scoring within each of five performance categories on tests for several grade levels and several subjects.  Progress toward Improvement Target is based on the amount of progress a school must make on its proficiency index every two years to reach the target for all state schools and districts by 2014.	Prior performance.	No Financial Rewards. Sanctions for low API and no growth; sanctions include assistance and reconstitution.  Publicly Available Report Cards.
Michigan	School. School District. <u>Subgroups:</u> None.	School grade is a single score for each school based on both test scores and 3 other factors. School grade is part of the Accreditation process for each school.	The Standards for Accreditation depend upon Measures of School Performance (indicators of engagement, indicators of instructional quality, and indicators of learning opportunities, which include attendance and dropout rates) and Measures of Student Achievement (achievement status, achievement change, and achievement growth). The six components will be combined (2/3 weight on achievement) into a letter grade. Schools with a grade of F are Unaccredited.	Prior performance.	No Financial Rewards. Sanctions for schools denied accreditation; sanctions include assistance.  Publicly Available Report Cards.
Minnesota	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	<i>No analysis; report card of test scores and other school data.</i>			

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Mississippi	School. School District. <u>Subgroups:</u> (1) Race (2) Special Education (3) LEP	Achievement Index is a single score (level) for each school (1 to 5) based on state test scores. Growth Model signifies whether the students in the school failed to grow, made expected growth, or exceeded expected growth.	Achievement Index is a single score (level) for each school based on the fraction of all students that achieve at each of several proficiency levels on state tests for several grade levels and subjects. Growth Model signifies whether the students in the school failed to grow, made expected growth, or exceeded expected growth by 10% on scaled scores on state tests for several grade levels and subjects. (Expected growth is estimated using a regression model that takes into account student proficiency levels.) Schools are placed in one of 5 School Performance Classifications based on Achievement and Growth.	Prior performance of students is controlled in the regression prediction model.	No Financial Rewards. No Sanctions. Publicly Available Report Cards.
Missouri	School. School District. <u>Subgroups:</u> None.	Missouri School Improvement Plan Accreditation, based on test scores (level and growth) and 2 other factors, is conducted for each school district.	The Accreditation process depends upon indicators of Resources, Processes, and Performance. Points are assigned for meeting resource requirements, meeting process requirements (e.g. instructional design, practice, school services), and meeting achievement requirements, which are based on proficient performance on the state exam and improvement on the state exam. School districts are then labeled Accredited, Provisionally Accredited, or Unaccredited.	Prior Performance.	No Financial Rewards. Sanctions for school districts denied accreditation; sanctions include assistance and reconstitution. Publicly Available Report Cards for schools.
Montana	None.	<i>No analysis; no published school-level report card of test scores and other school data.</i>			None.
Nebraska	School. School District. <u>Subgroups:</u> (1) Special Education (2) LEP	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Nevada	School. School District. <u>Subgroups:</u> None.	School designation is a single score (from 4 categories) for each school based on the level of state test scores, attendance, and % students tested.	School Designation is a single score (from 4 categories) for each school based on: (1) % students taking exam. (2) % of students scoring at various performance levels on state tests for several grades and several subjects. (3) Attendance rate.	None.	No Financial Rewards. Sanctions for schools labeled need for improvement; sanctions include assistance and reconstitution. Publicly Available Report Cards.
New Hampshire	School. School District. <u>Subgroups:</u> None.	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.
New Jersey	School. School District. <u>Subgroups:</u> (1) Special Education (2) LEP	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.
New Mexico	School. School District. <u>Subgroups:</u> None.	Overall School Rating is a single score (combining level and growth) for each school (one of four categories) based on state test scores, attendance, and dropout rate.	Overall School Rating is a single score (combining level and growth) for each school (one of four categories) based on state test scores, attendance, and dropout rate. The formula gives points for median (level) test scores for several grade levels and subjects, for growth in test scores, for low dropout rate, and for high attendance. The points are then translated into one of four rating categories.	Prior performance.	Rewards (financial) for high improving schools. Sanctions for schools with low overall school rating; sanctions include private management. Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
New York	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	Accountability Performance Index (API) is a score (level) calculated separately for English Language Arts and Math tests. Schools that fall below the state standard index are assigned an Adequate Yearly Progress (AYP) Target.	API scores range from 0-200 based on fraction of students scoring at or above proficiency levels 2 and 3 (on a scale of 1 to 4) on state tests.  Schools are required to meet the State Standard API (145 in 2001-02, 150 in 2002-03) score. Those that do not are required to reduce the difference between the school API and the State Standard by 15% per year over three years, for a total of 45% after three years. This is the AYP Target.	Prior performance for below standard schools.	No Financial Rewards. Sanctions for schools failing to meet AYP targets for two straight years are under review; sanctions include assistance, closure, and reconstitution.  Publicly Available Report Cards.
North Carolina	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	Performance Composite is a single score (level) for each school based on state test scores.  Expected Growth Composite signifies whether the students in the school made expected growth.  Exemplary Growth Composite signifies whether the students in the school exceeded expected growth.	Performance Composite is a single score (level; numeric) for each school based on the fraction of all students that achieve at the proficient level on state tests for several grade levels and subjects.  Expected Growth Composite signifies whether the students in the school made expected growth in scaled scores on state tests for several grade levels and subjects. (Expected growth is estimated using a regression model that takes into account student proficiency levels and the effect of regression to the mean.)  Expected Growth Composite signifies whether the students in the school exceeded expected growth in scaled scores by 10% on state tests for several grade levels and subjects.  Schools are placed in one of 6 Recognition Categories based on three composites above.	Prior performance of students is controlled in the regression prediction model.	Rewards (financial) for schools that earn high growth (\$1500 per teacher) or expected growth (\$750 per teacher).  Sanctions for school recognition category of Low-Performing; sanctions include assistance, reconstitution, student transfers, and private management.  Publicly Available Report Cards.
North Dakota	None.	No analysis; no published school-level report card of test scores and other school data.			None.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Ohio	School. School District. <u>Subgroups:</u> (1) Race	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.
Oklahoma	School School District <u>Subgroups:</u> (1) Special Education	Accountability Performance Index (API) is a single numeric score (level, from 0-1500) for each school based on test scores and 5 other factors. Progress of the API toward Growth Target for each school.	API scores range from 0-1500 based on: fraction of students scoring at each of 4 grades and several subjects (80% weight), attendance rate, graduation rate, dropout rate, ACT score and participation, Advanced Placement scores, and college remediation rate. Progress toward Growth Target is based on the amount of progress a school must make to reach the State's 10 year goal (API=1250) by 2014.	Prior performance.	Rewards (financial) for schools that meet or exceed growth targets. Sanctions for low API; sanctions include assistance, closure, reconstitution, student transfers, private management. Publicly Available Report Cards.
Oregon	School. School District. <u>Subgroups:</u> None.	Overall School Performance Rating is a single score (from 5 categories) for each school based on the level and growth of state test scores, attendance, dropout rate, and % students tested.	Overall School Performance Rating is a single score (from 5 categories) for each school based on: (1) % of students scoring at various performance levels on state tests for several grades and several subjects. (2) Attendance and dropout rates. (3) Improvement in % of students scoring at various performance levels. (4) Improvement in attendance and dropout rates. (5) % students taking exam.	Prior performance.	No Financial Rewards. Sanctions for schools labeled low or unacceptable; sanctions include assistance. Publicly Available Report Cards.



(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Pennsylvania	<p>School. School District.</p> <p><u>Subgroups:</u> None (although online reporting system does present subgroup data)</p>	<p>School Performance Funding (SPF) initiative rewards schools based on composite scaled scores (level) on state tests; growth on state tests, and growth on attendance rate.</p>	<p>School Performance Funding (SPF) initiative rewards schools based on any one or more of the following achievements:</p> <p>(1) Composite scaled score (level) on state tests for several grades and subjects. (2) Improvement in composite scaled score on state tests for several grades and subjects. (3) Improvement in attendance rate.</p>	<p>Prior performance.</p>	<p>Rewards (financial) for schools selected in SPF initiative. No Sanctions or Assistance (system does not identify low-performing schools). Publicly Available Report Cards.</p>
Rhode Island	<p>School. School District.</p> <p><u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP</p>	<p>Adequate Yearly Progress (AYP) is the progress toward goal of 100% student proficiency on state tests.</p>	<p>AYP is not based on a mathematical formula. Rather, schools are required to set their own three-year targets in terms of the % of students they will move out of the lowest levels of achievement (in 5 performance categories) into higher levels. Making AYP requires meeting these goals.  (On report cards, but not part of AYP, is a value-added measure that controls for socioeconomic differences in schools. Regression model controls for: (1) % free lunch; (2) % minority; (3) % with parents with some college; (4) % LEP; (5) % special ed in school.)</p>	<p>No Controls for AYP. Controls for report cards include: (1) % free lunch; (2) % minority; (3) % with parents with some college; (4) % LEP; (5) % special ed in school.</p>	<p>No Financial Rewards. Sanctions for not making AYP; sanctions include assistance, closure, and reconstitution. Publicly Available Report Cards.</p>

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
South Carolina	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP (5) Migrant	Absolute Performance Index (API) is a single score (level) for each school (from 0 to 5) based on state test scores. Improvement Index is the growth in API from the previous year.	API is a single score (level) from 0 to 5 for each school based on weighted average on state test scores for several grade levels and subjects (based on five possible performance categories). Improvement Index is calculated as the growth in school API score from the previous year to the current year (based on longitudinally matched student data). For each index above, score is translated into one of 5 rating categories. If improvement index for each subgroup exceeds the state index by 1 standard deviation, the improvement rating will increase by 1 level.	Prior performance.	Rewards (financial) for Palmetto Gold and Silver Awards based on both API and Improvement Index. Sanctions for schools with API ratings of Below Average or Unsatisfactory; sanctions include assistance, closure, and reconstitution. Publicly Available Report Cards.
South Dakota	School. School District. <u>Subgroups:</u> None.	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.
Tennessee	School. School District. <u>Subgroups:</u> None.	Multiple grades (A through F) are provided for test scores (level), test score value-added (growth), attendance, dropout rate, and promotion rate. There is no composite school rating provided.	Multiple grades (A through F) are provided for test scores by subject area in the form of median national percentile scores and accompanying letter grades. Separate scores are provided for each subject area. Separate grades are also provided for attendance, dropout, and promotion rates for each school. A regression prediction model is used to provide separate letter grades for value-added for each specific subject area.	Prior performance of students is controlled for in the regression prediction model.	Rewards (financial) for schools that meet achievement standards. Sanctions for schools failing to meet achievement standards; sanctions include assistance, closure, reconstitution, student transfers, private management, and withholding funds. Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Texas	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education	School rating is a single measure (level) for each school based on fraction of students on passing state test and on dropout rate. There is a similar district rating.  Comparable Improvement (CI) indicator measures growth from prior year for schools as compared to similar schools.	School rating is a single measure (level) for each school based on: the fraction students that achieve at the proficient level on state tests for several grade levels and subjects; and on the school dropout rate. All students as well as each of the following subgroups (African American, White, Hispanic, and Economically Disadvantaged).  Ratings (4 categories) are based on the lowest score for that school (i.e. a low score in one subject exam for one subgroup will move school rating to that level).	CI comparison group is 40 schools, matched according to ethnicity, economic disadvantage, Limited English proficiency, and student mobility.  Prior performance.	Rewards (financial, \$500-\$3,000) based on rating of Acceptable or above with gains on CI.  Sanctions for School Rating of Low-Performing; sanctions include assistance, closure, reconstitution, and student transfers.  Publicly Available Report Cards.
Utah	None.	<i>No analysis; no published school-level report card of test scores and other school data.</i>			None.
Vermont	School. School District. <u>Subgroups:</u> None.	Status Index is a single score (level) for each school (from 0 to 600) based on state test scores.  Growth Score is a measure of the progress toward growth target for each school.	Status Index is a single score (level) for each school based on test scores on several grade levels and subjects (range from 0-140).  Score is the weighted average based on the fraction of students scoring within each of five performance categories on state exams.  Growth Score is a ratio of the actual growth to the target growth; target growth is based on the amount of progress a school must make to reach the State's goal score of 500 by 2007-08. (based on longitudinally matched student data).	Prior performance.	No Financial Rewards.  Sanctions for schools based on very low status index or low status and low growth; sanctions include assistance, closure, and reconstitution.  Publicly Available Report Cards.
Virginia	School. School District. <u>Subgroups:</u> None.	Standards of Accreditation (SOA) evaluation, based on test scores, is conducted for each school.	SOA rating is based on student achievement on state tests for several grade levels and subjects. Schools must achieve state-determined pass rates to retain accreditation.	None.	No Financial Rewards.  Sanctions for schools labeled Accredited with Warning; sanctions include assistance.  Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Washington	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	<i>No analysis; report card of test scores and other school data only.</i>			Publicly Available Report Cards.
West Virginia	School. School District. <u>Subgroups:</u> None.	School Accreditation Status (5 levels), based on test scores (level), attendance, and dropout rate.	School Accreditation Status (5 levels), is based on test scores, attendance, and dropout rate. Schools are labeled Seriously Impaired Status if (any of the following exist): (1) 30% or fewer of students perform at or above 3 <sup>rd</sup> quartile on state tests. (2) Attendance rate at or below 80%. (3) Dropout rate at or above 9%. (4) Below full accreditation criteria in all 3 categories (test scores, attendance, dropout).	None.	No Financial Rewards. Sanctions for schools labeled Seriously Impaired Status; sanctions include assistance, reconstitution, and student transfers. Publicly Available Report Cards.
Wisconsin	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	Annual Review Criteria, based on test scores (level) and test score growth.	To satisfy the Annual Review Criteria, schools must meet one of the following two requirements: (1) Meet the minimum requirements for % proficient on state tests for several grade levels and subjects (2) Meet the required Continuous Progress Indicator (CPI) of growth; CPI measured by increase in % of proficient students and decrease in the % of below basic students	Prior performance.	No Financial Rewards. Sanctions for schools not meeting annual review criteria for two consecutive years; sanctions include assistance. Publicly Available Report Cards.

(1) State	(2) Assessment Unit	(3) Output Measures	(4) Output Calculations	(5) Controls or Inputs	(6) Consequences
Wyoming	School. School District. <u>Subgroups:</u> (1) Race (2) Poverty (3) Special Education (4) LEP	No analysis; report card of test scores and other school data only.			Publicly Available Report Cards.

**Notes on Data Sources:**

- For columns 3,4,5, *Education Week Quality Counts 2003*\* Standards and Accountability Table (beginning on page 84) was employed to determine whether the state assigns ratings to all schools or identifies low-performing schools. There are 29 such states with some type of rating system. In addition, there were three states identified as not having a rating system, but as states that offer financial rewards to high performing or improved schools (Georgia, Indiana, and Pennsylvania). These three states were investigated in more detail to determine whether or not an accountability system is in place. In Georgia, the rewards referred to a Pay for Performance plan, which involve voluntary participation and application by schools; thus, it was determined that Georgia does not have an accountability rating system. In Indiana, the rewards reported in *Education Week Quality Counts 2003* referred to an aspect of the accountability plan that is to be implemented in the future, but is not currently in place; thus, it was determined that Indiana does not have an accountability rating system. In Pennsylvania, the rewards reported in *Education Week Quality Counts 2003* referred to the School Performance Funding (SPF) Initiative. The SPF initiative ranks all schools in the state based on achievement and improvement on state tests and attendance rate; thus, it was determined that Pennsylvania does have an accountability rating system. Therefore, 30 states are classified as having an Accountability Rating System.
- For these 30 states, information for columns 3,4,5 was gathered from the individual State Department of Education websites. This information was augmented by the survey of state assessment and accountability systems conducted by the Consortium for Policy Research in Education (CPRE) in 2000. ([http://www.cpre.org/Publications/Publications\\_Accountability.htm](http://www.cpre.org/Publications/Publications_Accountability.htm)).
- For those states with no rating system identified, columns 3,4,5 indicate that there is no rating system.
- All subgroup information highlighted in Column 2 was taken directly from *Education Week Quality Counts 2003*, (January 2003), Standards and Accountability Table (beginning on page 84).
- All information related to School Report Cards, Assistance, and Sanctions in Column 6 was taken from *Education Week Quality Counts 2003*, (January 2003), Standards and Accountability Table (beginning on page 84).
- Information related to financial rewards for schools was gathered from the individual State Department of Education websites.

\* Quality Counts 2003: If I Can't Learn From You, (Special Issue: Annual Report published in collaboration with the Pew Charitable Trusts), *Education Week*, Volume 22, Number 17, January 9, 2003. (web address: <http://www.edweek.org/sreports/QC03/>)

## **VI. CONCLUSIONS: CALIFORNIA'S ASSESSMENT PROGRAM**

This is the national context in which California's assessment program can be appraised. Decisions about using one or multiple output measures, levels or changes in levels, standards or continuous measures, controls incorporated into prior performance or specified inputs or no controls, and the magnitude and nature of sanctions/reward consequences are the product of negotiations among multiple stakeholders. Legislatures, courts, Governors, State Departments of Education, teacher unions, Superintendents of schools, principals, and parents all weigh in with varying degrees of strength. Many of these draw on the extensive relevant research; many do not.

California's system ranks high on the merits of its assessment plan in relation to the rest of the country. The characteristics of its plan meet many of the standards emerging from the most rigorous research in the field, placing it in the top group of state educational systems.

- There is regular, impartial, quantitative measurement of the learning of cognitive skills in every school.
- Both levels and changes in levels are incorporated into the assessment measure.
- The assessment index uses both a modified continuous measure, and a threshold measure (distance from the annual growth target).
- There is focus on the outcomes for special groups of interest—race, poverty, LEP, and Special Education.
- Controls are incorporated into the assessment measure. Prior performance and comparisons to similar schools are significant input measures.

- Output measures trigger sanctions and rewards. The data are not used with excessive precision. The decile bands create reasonable latitude.
- There are incentive consequences to the assessment calculations--both rewards and sanctions.

As with any existing state program, there are additions to the assessment system that would improve California's assessment system: (1) It would be desirable to add multiple measures of output. As was stated above in Section III, though cognitive skills are a necessary and significant measure of learning, supplementary measures of dropouts and postsecondary performance would be important additions. (2) School averages, used in the calculation of the Academic Performance Index, are, of course, derived from individual school data at some point. Ideally, instead of using averages of classes to get school averages and averages of schools to get school district averages, cohorts of identical students should be tracked.

The search for school-controlled inputs that significantly affect outputs should be part of the regular research activity of all State Departments of Education. When an input is empirically determined to be significant, then its quantity and allocation should be on the state's education agenda. Several of plaintiff's expert papers use small-scale, non-random surveys and simple, uncontrolled correlations between one output and one input to argue for focusing on particular inputs. They do not provide any scientific basis for action.

California's assessment system incorporates the generally recognized major inputs into student learning—SES and prior school inputs. Hard evidence should precede the use of other inputs. The notion that California's system has a "single-minded focus on outcomes", as one of the expert papers states explicitly, is simply wrong. The use of the "similar schools" comparisons and the use of growth measures, by definition, define their extensive use of inputs.

Improvements in California's system of accountability are likely to be made over time, since assessment schemes are dynamic -- they are improved as they are used and understood better. But their methodology with inputs should not be on the current in-need-of-repair list.



## Footnotes

<sup>1</sup> Among others, see Hanushek, Eric A. (1986). "The Economics of Schooling: Production and Efficiency in the Public Schools", Journal of Economic Literature 24 (3); Hanushek, Eric A. et al. (1994). Making Schools Work: Improving Performance and Controlling Costs. Washington, DC: Brookings Institution; Summers, Anita A. and Barbara L. Wolfe (1977). "Do Schools Make a Difference?" American Economic Review 67 (4); Odden, Allan and Kelley, Caroline (1997). Paying Teachers for What They Know and Do. Thousand Oaks, CA: Corwin Press, Inc.; Committee for Economic Development (1994). Putting Learning First: Governing and Managing the Schools for High Achievement. New York, NY: CED.

<sup>2</sup> Hanushek, Eric A. (2002). "The Importance of School Quality", Unpublished Paper. Stanford, CA: Hoover Institution.

<sup>3</sup> Hoxby, Caroline M. (Winter 2002/2003). "Program Report: Economics of Education", NBER Reporter. Cambridge, MA: National Bureau of Economic Research.

<sup>4</sup> Most of the list of objections to testing and responses is taken from Summers, Anita A. (2002). "Expert Measures", Education Next, 2 (2). Stanford, CA: Hoover Institution.

<sup>5</sup> Hanushek, Eric A. and Raymond, Margaret E. "Lessons and Limits of State Accountability Systems", Paper for Conference on Taking Account of Accountability: Assessing Policy and Politics, Harvard University, June 9-11, 2002.

<sup>6</sup> Raymond, Margaret E. and Hanushek, Eric A. "High-Stakes Research", forthcoming in Education Next, Summer 2003. Stanford, CA: Hoover Institution.

<sup>7</sup> Ibid.

<sup>8</sup> Smith, M.S., Scoll, B.W., Link, J. (1995). Research-Based School Reform: The Clinton Administration's Agenda", Improving the Performance of America's Schools, National Research Council. Washington, D.C.: National Academy Press; Hanushek, Eric A., "Outcomes, Costs, and Incentives in Schools", in same volume. Also see footnote 1 references.

<sup>9</sup> Summers, A.A. and Wolfe, B.L. (1977), Ibid.

## Anita A. Summers

Anita A. Summers is Professor Emeritus of Public Policy and Management in the Wharton School of the University of Pennsylvania. She has secondary appointments in the Graduate School of Education and the Real Estate Department. She is a Senior Research Fellow in the Wharton Real Estate Center, and member of the Brookings Institution project on Metropolitan Solutions to Urban and Regional Problems.

Professor Summers did her graduate work at the University of Chicago and Columbia University, and received an Honorary Degree from Hunter College in 1995. At the University of Pennsylvania, she was Chair of the Public Policy and Management from 1983-1988. She has been a Visiting Professor at the Hebrew University of Jerusalem, Temple University and the University of Hawaii, she was a Resident Scholar at the Rockefeller Foundation Study Center in Bellagio, Italy in 1986; and was Research Officer and Head, Urban Research Section of the Federal Reserve Bank of Philadelphia from 1975-1979. She was on the Provost's Academic Planning and Budget Committee from 1984-1990, Member of the Law School and Wharton School Dean Search Committees, and Chair of the Research Policy Committee at the University of Pennsylvania. She is currently Ombudsman for the University of Pennsylvania.

Professor Summers' research work has been in two major areas--education efficiency and urban economic development and finance. She has published extensively in both. Her most recent works, following the publication of three books on the Philadelphia metropolitan area are Urban Change in the U.S. and Western Europe: Comparative Analysis and Policy (edited with Paul Cheshire and Lanfranco Senn, published by the Urban Institute Press in 1993, second edition in 1999), Schooling and Labor Market Performance (with Amy W. Johnson and David Crawford, Economics of Education Review, 1997), contributor to Making Schools Work (Eric A. Hanushek, et al, The Brookings Institution, 1994), and "Accountability Gains: Expert Measures", Education Next, 2002. She recently completed working on a project, funded by the Pew Charitable Trusts, on Regionalization: The Feasible Options, has co-authored the first Brookings Institution's Policy Briefs- Paying for the Poor: A New Strategy for Handling the Urban Burden, and is co-authoring a chapter in the forthcoming Brookings volume, Forging Metropolitan Solutions to Urban and Regional Problems.

Professor Summers was on the Board of Directors of The William Penn Foundation, and is currently Chair of the Board of Directors of Mathematica Policy Research, Inc. She was a member of the National Panel on the Economics of Educational Reform, the Tax Policy Advisory Committee for the City of Philadelphia, and the Metropolis Project of the American Academy of Arts and Sciences; and is currently part of the Brookings Institution Urban and Metropolitan Policy Center Project. She has consulted with several local school districts on the design and implementation of assessment schools, and has been an expert witness on school finance for a number of states and school districts.

### **Gary W. Ritter, Ph.D.**

Gary W. Ritter is an Assistant Professor of Education and Public Policy at the University of Arkansas. Dr. Ritter earned his Ph.D. degree in Education Policy from the Graduate School of Education of the University of Pennsylvania. Prior to that, he earned an M.A. in Public Policy in Management from the Wharton School of the University of Pennsylvania, and an M.A. in Social Policy from the University of Manchester in England.

Dr. Ritter has been on faculty at the University of Arkansas since January 2000. Currently, he is Associate Director of the University's inter-disciplinary Public Policy Ph.D. Program. Dr. Ritter has presented approximately thirty papers at national research conferences including the Association of Public Policy and Management (APPAM) and the American Educational Research Association (AERA) and has produced 10 research articles in peer-reviewed journals, including the *Journal of Education Finance, Educational Evaluation and Policy Analysis*, and the *Georgetown Public Policy Review*. Currently, he teaches courses in Education Policy, Program Evaluation, and Analysis of Large Scale Databases. He advises doctoral students and conducts research in school accountability, school finance, and program evaluation.

From 1997 through 2001, Dr. Ritter served as a research assistant to Professor Anita Summers, providing research and statistical support, in her work for legal counsel representing the City of Philadelphia in a school-funding lawsuit against the Commonwealth of Pennsylvania. In addition, Dr. Ritter served as an expert witness from 1998-2001 conducting statistical analysis and producing written legal testimony for legal counsel representing seventeen low-wealth New Jersey school districts in a school-funding lawsuit against the State of New Jersey. He delivered 3 days of testimony before the Administrative Law Judge in Trenton, New Jersey in September 2001. Also in 2001, Dr. Ritter conducted statistical analysis and preparation for legal testimony for legal counsel representing forty-two middle-wealth New Jersey school districts in a school-funding lawsuit against the State of New Jersey.