

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

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

Attorneys for Defendant State of California

SUPERIOR COURT OF THE STATE OF CALIFORNIA
CITY AND COUNTY OF SAN FRANCISCO

ELIEZER WILLIAMS, et al.,) Case No. 312 236
))
) Plaintiffs,) Date Action Filed: May 17, 2000
))
) vs.) .
))
STATE OF CALIFORNIA, DELAINE)
EASTIN, State Superintendent)
Of Public Instruction, STATE)
DEPARTMENT OF EDUCATION, STATE)
BOARD OF EDUCATION,)
))
) Defendants.)
))
STATE OF CALIFORNIA)
))
) Cross-Complainant,)
))
) vs.)
SAN FRANCISCO UNIFIED SCHOOL)
DISTRICT, et al.,)
))
) Cross-Defendants.)
))

EXPERT WITNESS DECLARATION RE SUSAN E. PHILLIPS, Ph.D.

1 I, Paul B. Salvaty, declare as follows:

2

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

6

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

11

12 3. Dr. Phillips has agreed to testify at trial.

13

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

18

19 5. Dr. Phillips' fee for providing deposition testimony,
20 consulting with the State, conducting research and other
21 activities undertaken in preparation of the attached report is
22 \$300 per hour.

23


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

28

1 7. Attached hereto as Exhibit B and incorporated herein by
2 reference is Dr. Phillips' 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. Phillips is expected to give at trial.
6 Dr. Phillips provides a comprehensive overview of California's
7 API, including how it was created, how it's been implemented, and
8 how it will be refined and improved in the next few years.
9 Phillips also provides a rebuttal to the critique of the API
10 presented in plaintiffs' expert reports, with specific focus on
11 the criticisms by plaintiffs' expert Michael Russell. The
12 foregoing statements are only a general summary of the issues and
13 conclusions discussed and documented more fully in Dr. Phillips'
14 expert report.

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

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

21
22 
23 _____
24 Paul B. Salvaty

VITA

S. E. PHILLIPS

Box 384

W. Paducah, KY 42086

e-mail: PhillipsSE@earthlink.net

(520) 465-6623

EDUCATION

<u>Institution</u>	<u>Location</u>	<u>Years</u>	<u>Degree</u>	<u>Date</u>
Cooley Law School	Lansing, MI	1988-90	JD	5 / 90
University of Iowa	Iowa City, IA	1976-81	PhD	12 / 81
University of Iowa	Iowa City, IA	1974-75	MA	12 / 75
University of Iowa	Iowa City, IA	1972-74	BA	5 / 74
Iowa State University	Ames, IA	1970-72		

EXPERIENCE

Independent Consultant	Assessment Law	2000-present
Professor	Michigan State University	1994-2000
Adjunct Professor	T. M. Cooley Law School	Fall 1991
Associate Professor	Michigan State University	1987-1994
Psychometric Consultant	National Computer Systems	1983-present
Visiting Scholar	The Psychological Corporation	Winter 1987
Assistant Professor	Michigan State University	1982-1987
Post-Doctoral Associate	University of Iowa	1982
Statistical/Programming Consultant	Westinghouse DataScore	1982
Teaching Assistant (statistics)	University of Iowa	1981
Measurement Specialist Co.		Riverside Publishing 1979-81
Tutoring and Consulting with Business and Computer Majors taking Statistics Courses		1979-81
Program Specialist	American College Testing Program (ACT)	1978-79

Director Project Metric	ACT	1977-78
Research Associate	ACT	1976-77
Economics Instructor	Coe College Cedar Rapids, IA	1976-77
Mathematics Instructor	Mt. Mercy College Cedar Rapids, IA	1975-76
Substitute Teaching	Iowa City Community Schools	1974-75

RESEARCH AND SCHOLARSHIP

Publications -- Legal

Phillips, S.E. *G.I. Forum v. TEA: Psychometric Evidence*, APPLIED MEASUREMENT IN EDUCATION (in press).

Phillips, S.E. *Legal Issues Affecting Large-Scale Testing Programs*, invited chapter in Tindal, G. & Haladyna, T. (Eds.), *Large-Scale Assessment Programs for ALL Students: Development, Implementation, and Analysis* (in press).

Phillips, S.E. *Legal Issues and Considerations in Standard Setting for K-12 Programs*, invited chapter in Cizek, G.J. (Ed.), *Standard Setting: Concepts, Methods, and Perspectives* (in press).

Phillips, S.E. *Assessment Accommodations*, Fall 1997(3) DETROIT COLLEGE OF LAW AT MICHIGAN STATE UNIVERSITY L. REV. 917 (1997).

Phillips, S.E. *Test Disclosure Policies for Large-Scale High-Stakes Assessments*, (manuscript in preparation).

Phillips, S.E. *Legal Criteria For School Accountability Assessments*, (manuscript in preparation).

Phillips, S.E. *Assessment Accommodations For LEP Students*, (manuscript in preparation).

Millman, J. & Phillips, S.E. *Performance-Based Measures of Lawyering Skills: An Alternative*, (manuscript in preparation).

Phillips, S.E. *Legal Criteria For Performance Assessments*, (under journal review).

Phillips, S.E. *The First and Fourteenth Amendments on College Campuses: Free Speech Versus Politically Correct Speech*, (under revision).

- Phillips, S.E. *Legal Defensibility of Standards: Issues and Policy Perspectives Part II*, 98(4) ARTS EDUCATION POLICY REVIEW, March/April 1997, reprinted from EMIP, 1996.
- Phillips, S.E. *Legal Defensibility of Standards: Issues and Policy Perspectives Part I*, 98(3) ARTS EDUCATION POLICY REVIEW, January/February 1997, reprinted from EMIP, 1996.
- Phillips, S.E. *Evaluation of Fairfield Test Data Re Alleged Tampering*, report commissioned by the Fairfield, CT School District Board of Education, October 1996.
- Phillips, S.E. *Legal Defensibility of Standards*, Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments, Washington, DC, (INVITED), September 1996.
- Phillips, S.E. *Legal Defensibility of Standards: Issues & Policy Perspectives*, 15(2) EDUC. MEAS: ISSUES & PRACTICE 5, Summer 1996.
- Phillips, S.E. *All Students, Same Test, Same Standards: What the New Title I Legislation Will Mean for the Educational Assessment of Special Education Students*. North Central Regional Educational Laboratory, Oak Park, Ill. (INVITED, 1996) (with editorial assistance from Mary Atteberry).
- Millman, J., Phillips, S.E., & Weil, R. *Guidelines For Measuring Lawyering Skills on the New York Bar Examination*, Report prepared for the New York Board of Bar Examiners, November 1995.
- Phillips, S.E. *High-Stakes Testing Accommodations: Validity Versus Disabled Rights*, THE BAR EXAMINER 8, August 1995.
- Phillips, S.E. *Test Disclosure Policies for the Mississippi FLE Program*, Report prepared for the Mississippi Department of Education, August 1995.
- Hambleton, R.K., Jaeger, R.M., Koretz, D., Linn, R.L., Millman, J., Phillips, S.E., *Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991-1994 Final Report*, A Report Prepared for the Office of Educational Accountability, Kentucky General Assembly, June 20, 1995 (sole author 55-page Legal Appendix; first author 29-page chapter on equating).
- Phillips, S.E. *Legal Implications of High Stakes Assessment: What States Should Know*, Handbook commissioned and published by the North Central Regional Educational Laboratory, November, 1994.
- Phillips, S.E. *High-Stakes Testing Accommodations: Validity Versus Disabled Rights*, 7(2) APPLIED MEAS. IN EDUC. 93 (1994).
- Phillips, S.E. *Testing Condition Accommodations For Disabled Students*, 80 ED. LAW REP. 9 (March 25, 1993).

Phillips, S.E. *Legal Issues in Performance Assessment*, 79 ED. LAW REP. 709 (March 11, 1993).

Phillips, S.E. *Extending Teacher Licensure Testing: Have the Courts Applied the Wrong Validity Standard?* 8(3) T. M. COOLEY L. REV. 513 (1991).

Phillips, S.E. *Diploma Sanction Tests Revisited: New Problems From Old Solutions*, 20(2) J. LAW & EDUC. 175 (1991).

Phillips, S.E. *The Golden Rule Remedy for Disparate Impact of Standardized Testing: Progress or Regress?* 63 ED. LAW REP. 383 (Dec. 20, 1990).

Phillips, S.E. *High School Grade Reductions for Absenteeism: Incentive or Curse?* 6(1) T. M. COOLEY L. REV. 129 (1989).

Publications -- Psychometric

Clarizio, H.F., Payette, K. A., & Phillips, S.E. *A Comparison of Methods for Determining Learning Disabilities: Effects on Racial Representation*, (in press).

Berk, R., Phillips, S.E. & Poggio, J. *Recommendations Based on the TAAS-TASP Equivalence Meeting*, Report prepared for the Texas Education Agency, February 1996.

Payette, K.A., Clarizio, H.F., Phillips, S.E., & Bennett, D.E. *The Effects of Simple and Regressed Discrepancy Models and Cutoffs on Severe Discrepancy Determination*, 32 PSYCHOLOGY IN THE SCHOOLS 93, April 1995.

Clarizio, H.F. & Phillips, S.E. *A Comparison of Severe Discrepancy Formulae: Implications For Policy Consultation*, 3(1) J. EDUC. & PSYCHOLOGICAL CONSULTATION 55 (1992).

Mehrens, W.A., Phillips, S.E., & Schram, C. *Survey of Statewide Test Security Practices*, 12(4) EDUC. MEAS.: ISSUES & PRACTICE 5 (1993).

Fugate, D., Clarizio, H.F. & Phillips, S.E. *Referral to Placement Ratio: A Finding in Need of Replication?*, 26 J. LEARN. DISABILITIES 413 (1993).

Phillips, S.E. & Clarizio, H.F. *Conflicting Growth Expectations Cannot Both Be Real: A Rejoinder to Yen*, 7(4) EDUC. MEAS.: ISSUES & PRACTICE 18 (1989).

Mehrens, W.A. & Phillips, S.E. *Using College GPA and Test Scores in Teacher Licensure Decisions: Conjunctive Versus Compensatory Models*, 2(4) APPLIED MEAS. IN EDUC. 277 (1989).

- Clarizio, H.F. & Phillips, S.E. *Defining Severe Discrepancy in the Diagnosis of Learning Disabilities: A Comparison of Methods*, 27 J. OF SCHOOL PSYCHOLOGY 383 (1989).
- Phillips, S.E. & Mehrens, W.A. *Effects of Curricular Differences on Achievement Test Data at Item and Objective Levels*, 1(1) APPLIED MEAS. IN EDUC. 33 (1988).
- Phillips, S.E. & Clarizio, H.F. *Some Limitations of Standard Scores in Diagnosing Learning Disabilities*, 7(1) EDUC. MEAS.: ISSUES AND PRACTICE 8 (1988).
- Mehrens, W.A., McLarty, J.R., Rakow, E.A., & Phillips, S.E. *Fiscal Viability in Career-Ladder Decisions: An Empirical Investigation*, 2 J. OF PERSONNEL EVAL. IN EDUC. 103 (1988).
- Mehrens, W.A. & Phillips, S.E. *Sensitivity of Item Difficulties to Curricular Validity*, 24 J. OF EDUC. MEAS. 357 (1987).
- Phillips, S.E. & Mehrens, W.A. *Curricular Differences and Unidimensionality of Achievement Test Data: An Exploratory Analysis*, 24 J. OF EDUC. MEAS. 1 (1987).
- Lehmann, I.J. & Phillips, S.E. *Teacher Competency Testing Programs: A National Survey*, 6(1) EDUC. MEAS.: ISSUES AND PRACTICE 14 (1987).
- Lehmann, I.J. & Phillips, S.E. *Teacher Competency Testing in the United States*, ERIC MONOGRAPH SERIES (1987).
- Clarizio, H.F. & Phillips, S.E. *Some Limitations of Standard Scores in Diagnosing Learning Disabilities: A Critique*. 23 PSYCH. IN THE SCHOOLS 381 (1986).
- Mehrens, W.A. & Phillips, S.E. *Detecting Impacts of Curricular Differences in Achievement Test Data*, 23 J. OF EDUC. MEAS. 185 (1986).
- Clarizio, H.F. & Phillips, S.E. *Sex Bias in the Diagnosis of Learning Disabled Students*, 23 PSYCH. IN THE SCHOOLS 44 (1986).
- Phillips, S.E. *Deletion of Misfitting Persons When Vertically Equating Achievement Tests Via the Rasch Model*, 23 J. OF EDUC. MEAS. 107 (1986).
- Phillips, S.E. *Quantifying Equating Errors With Item Response Theory Methods*, 9 APPLIED PSYCH. MEAS. 59 (1985).
- Phillips, S.E. *Comparison of Equipercentile and Item Response Theory Equating When the Scaling Test Method is Applied to a Multilevel Achievement Battery*, 7 APPLIED PSYCH. MEAS. 267 (1983).

Reviews

Phillips, S.E. *Review of The Iowa Tests of Educational Development (ITED)*, BUROS TENTH MENTAL MEASUREMENTS YEARBOOK 398 (1989).

Phillips, S.E. *Review of Thinking About My School*, BUROS TENTH MENTAL MEASUREMENTS YEARBOOK 863 (1989).

Phillips, S.E. *Review of The Basics of Item Response Theory by Frank B. Baker*, 23 J. OF EDUC. MEAS. 267 (1986).

Phillips, S.E. *Review of ACER Paragraph Reading Test*, BUROS NINTH MENTAL MEASUREMENTS YEARBOOK 21 (1985).

Phillips, S.E. *Review of Test of Enquiry Skills*, BUROS NINTH MENTAL MEASUREMENTS YEARBOOK 1570 (1985).

Phillips, S.E. *Review of Test of English as a Foreign Language (TOEFL)*. In Keyser, D.J. & Sweetland, R.C. TEST CRITIQUES VOLUME III. Kansas City: Test Corporation of America 655 (1985).

Phillips, S.E. *Review of Sequential Tests of Educational Progress (STEP III)*. In Keyser, D.J. & Sweetland, R.C. TEST CRITIQUES VOLUME I. Kansas City: Test Corporation of America 578 (1984).

Phillips, S.E. *Review of Applications of Item Response Theory by R.K. Hambleton Ed.*, 3 EDUC. MEAS.: ISSUES & PRACTICE, 27 (1984).

Newsletter Columns

Phillips, S.E. *School Assignments and Teacher Testing*, 5(4) NCME Newsletter 2, December 1999.

Phillips, S.E. *New Wrinkles in Performance Assessment*, 5(3) NCME Newsletter 2, July 1999.

Phillips, S.E. *Response to Freedman re Oklahoma Grading Case*, 5(2) NCME Newsletter 2, April 1999.

Phillips, S.E. *Ohio Update & New Grading Cases*, 5(1) NCME Newsletter 2, January 1999.

Phillips, S.E. *Selective Admissions in K-12 Programs*, 4(4) NCME Newsletter 2, November 1998.

Phillips, S.E. *Public Disclosure of State Graduation Tests*, 4(3) NCME Newsletter 2, July 1998.

Phillips, S.E. *Calculator Accommodations*, 4(2) NCME Newsletter 2, April 1998.

- Phillips, S.E. *High Stakes Assessment Raises Several Policy Issues*, 4(2) PERSPECTIVES 8, Winter 1998.
- Phillips, S.E. *Assessing Visually Impaired Examinees*, 4(1) NCME Newsletter 2, January 1998.
- Phillips, S.E. *Opportunity For Success*, 3(4) NCME Newsletter 2, November 1997.
- Phillips, S.E. *Assessment Accommodations*, AERA Division I NEWSLETTER (in press).
- Phillips, S.E. *Standards & Grading for Disabled Students – Part II*, NCME NEWSLETTER (in press).
- Phillips, S.E. *Standards & Grading for Disabled Students – Part I*, 4(1) NCME NEWSLETTER 2 April 1997.
- Phillips, S.E. *Policies For Public Review of Secure Assessments*, 3(4) NCME NEWSLETTER 2, April 1996.
- Phillips, S.E. *Parental Rights to View Statewide Assessments*, 3(2) NCME NEWSLETTER 2, November 1995.
- Phillips, S.E. *Content Challenges to Assessment Programs*, 2(3) NCME NEWSLETTER 2, February 1994.
- Phillips, S.E. *Update on Testing Accommodations, Part II*, 2(2) NCME NEWSLETTER 2, October 1993.
- Phillips, S.E. *Update on Testing Accommodations, Part I*, 2(1) NCME NEWSLETTER 2, June 1993.
- Phillips, S.E. *Introduction of the Legal Corner Column*, 1(2) NCME NEWSLETTER 2, April 1992.

Presentations

- Phillips, S.E. et al. *GI Forum v. TEA: A Challenge to the Texas Graduation Test*, Organized Symposium Accepted for National Conference on Large-Scale Assessment, Snow Bird, UT, June 2000.
- Phillips, S.E. *Reaction to More of Miss Marple's Measurement Moments*, Symposium Accepted for National Conference on Large-Scale Assessment, Snow Bird, UT, June 2000.
- Phillips, S.E. *Reaction to Equity Issues in Large-Scale High-Stakes Accountability Assessments: Some Perspectives from the Trenches*, Symposium Accepted for National Conference on Large-Scale Assessment, Snow Bird, UT, June 2000.

- Phillips, S.E. *Detecting Inappropriate Test-Preparation in the Classroom: Misdemeanor Offenses*, in Symposium on Instructionally Corrupt Test Preparation Accepted for AERA annual meeting, New Orleans, April 2000.
- Phillips, S.E. *Measurement Perspectives on an Oklahoma Grading Case*, in Symposium on Oklahoma Case Accepted for NCME annual meeting, New Orleans, April 2000.
- Phillips, S.E. *Access, Test Accommodations and Opportunity to Learn Issues*, High School Exit Examination Standards Panel Meeting, Sacramento, CA, November 18, 1999.
- Phillips, S.E. *Legal Issues and Accountability*, Delaware Chief School Officers Association Meeting, Dover, DE, November 4, 1999.
- Phillips, S.E. *The Stone Soup of Accommodations*, Keynote Speaker, Fall All City Special Education Conference, Chicago Public Schools, Chicago, IL, October 29, 1999.
- Phillips, S.E. *Needs, Wants, Access, Success: The Stone Soup of Accommodations*, National Conference on Large-scale Assessment, Snow Bird, UT, June 1999.
- Schafer, W.D., Rosenberger, K., Cruse, K., Phillips, S.E. *Miss Marple Meets Measurement: Security Investigation Models in Maryland and Texas*, National Conference on Large-scale Assessment, Snow Bird, UT, June 1999.
- Phillips, S.E. *Legal Issues in High-stakes Assessment*, HBEM Conference for District Policymakers, San Diego, CA, January 1999.
- Phillips, S.E., Thurlow, M., & Beck, M. *LEP Modifications: Why, What and When*, National Conference on Large-scale Assessment, Colorado Springs, June 14, 1998.
- Phillips, S.E. *Strategies for Evaluating School-Level Test Tampering*, National Conference on Large-scale Assessment, Colorado Springs, June 15, 1998.
- Phillips, S.E. *Update on Issues in Standards Development*, AERA Division D annual meeting featured symposium, Chicago, March 1997.
- Phillips, S.E. *Challenges in the Development of State Assessment Programs that Support Education Reform: Legal Considerations*, NCME annual meeting symposium, Chicago, March 1997.

- Phillips, S.E. *Legal Implications of Large-scale, High-stakes Assessment*, Second Annual Mississippi Assessment Symposium, Jackson, MS, February 1997.
- Phillips, S.E. *Legal/Policy Issues in Standard Setting for Large-scale Performance Assessments: Lessons Learned*, National Conference on Large-scale Assessment, Phoenix, June 25, 1996.
- Phillips, S.E. *Legal Issues and Assessment*, Virginia Association of Test Directors Annual Assessment Conference, Richmond, May 10, 1996.
- Phillips, S.E. *Faculty Evaluation of Students*, William Mitchell College of Law faculty retreat, Minneapolis, May 1, 1996.
- Phillips, S.E. *The Golden Rule Remedy for Disparate Impact of Standardized Testing*, National Conference of Bar Examiners Seminar on Bar Admissions, Chicago, April 20, 1996.
- Phillips, S.E. *Testing Under the ADA*, National Conference of Bar Examiners Seminar on Bar Admissions, Chicago, April 19, 1996.
- Phillips, S.E. *Accommodations for Assessment in Michigan*, Michigan School Testing Conference, February 27-28, 1996.
- Phillips, S.E. *High Stakes Accountability in Student Assessment*, North Carolina School Boards Association Fourth Annual Law Conference, Raleigh, February 22-23, 1996.
- Phillips, S.E. *Test Security in a High-Stakes Environment*, First Annual Mississippi Assessment Symposium, Jackson, February 8-9, 1996.
- Millman, J. & Phillips, S.E. *Alternative Item Formats For Measuring Lawyering Skills*, New York Board of Bar Examiners Meeting, Syracuse, November 3, 1995.
- Phillips, S.E. *Testing Under the ADA*, American Bar Association Conference, Chicago, August 6, 1995.
- Phillips, S.E. *Assessment Accommodations: Legal Perspectives & Policy Implications*, National Conference on Large-Scale Assessment, Phoenix, June 19-21, 1995.
- Phillips, S.E. *Legal Issues in Adaptation/Inclusion of IEP & LEP Students*, Michigan School Testing Conference, February 21-22, 1995.
- Phillips, S.E. *Legal Implications of High-stakes Assessment*, BELL Conference on Education Standards and Accountability, Minneapolis, MN, February 20, 1995 (INVITED).

- Phillips, S.E. *Legal Defensibility of Standards: Issues & Policy Perspectives*, NAGB/NCES Joint Conference on Standard Setting for Large-scale Assessments, Washington, DC, October 5-7, 1994.
- Phillips, S.E. *Legal & Political Issues Surrounding Performance Assessment*, National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Conference, Los Angeles, CA, September 11-13, 1994.
- Phillips, S.E. *Testing Accommodations: Validity Versus Disabled Rights*, Michigan School Testing Conference, February 1994.
- Phillips, S.E. *Legal Defensibility of Performance Assessments*, Wisconsin Outcomes Based Education Conference, Eau Claire, WI, July 1993.
- Phillips, S.E. *Legal Issues in High-Stakes Assessment*, Education Commission of The States Annual Conference, Albuquerque, NM, June 1993.
- Phillips, S.E. *Legally Defensible High-Stakes Assessments*, Colorado Assessment Conference, Breckenridge, CO, June 1993.
- Phillips, S.E. *Testing the Disabled: What is "Reasonable Accommodation"?* Michigan School Testing Conference, Ann Arbor, MI, February 1993.
- Phillips, S.E. *Legal Standards For Performance Assessments*, accepted for NOLPE Annual Convention, Phoenix, AZ, November 1992.
- Phillips, S.E. *Developing Legally Defensible Performance Assessments*, Education Commission of the States Annual Conference, Boulder CO, June 1992.
- Phillips, S.E. *Testing Accommodations for Handicapped Students*, Educ. Law SIG, AERA annual meeting, San Francisco, April 1992.
- Phillips, S.E. *Legal Aspects of Performance Assessment*, NCME annual meeting, San Francisco, April 1992.
- Mehrens, W.A., Phillips, S.E., & Schram, C. *Survey of Statewide Test Security Practices*, NCME annual meeting, San Francisco, April 1992.
- Phillips, S.E. *The Clash Between the First and Fourteenth Amendments on College Campuses: Free speech versus Discrimination*, NOLPE Annual Convention, Orlando, November 1991.
- Phillips, S.E. *Legal Issues in High Stakes Testing*, Education Commission of the States Annual Conference, Breckenridge, CO, June 1991.

- Phillips, S.E. *Legal Origins of Teaching the Test: The Debra P. Case*, NCME Annual Meeting Symposium I organized on *Implementation of Statewide Test Security Policies*, Chicago, April 1991.
- Phillips, S.E. *Legal Issues in the Reform of Teacher Testing*, AERA Annual Meeting, Chicago, April 1991.
- Phillips, S.E. *The Golden Rule Remedy for Disparate Impact of Standardized Testing*, NOLPE Annual Convention, San Antonio, November 1990.
- Phillips, S.E. *Diploma Sanction Testing Revisited*, MWERA Annual Meeting, Chicago, October 1990.
- Phillips, S.E. & Dutcher, P. *Equating the New MEAP Reading Test*. Michigan Educational Research Association Meeting, January 1990.
- Phillips, S.E. & Dutcher, P. *A Review of the New MEAP Reading Test*. Michigan School Testing Conference, February 1990.
- Phillips, S.E. *Comparison of Thurstonian & Rasch Methods of Equating Objective and Essay Writing Assessments*. NCME Annual Meeting, New Orleans, April 1988.
- Phillips, S.E. & Mehrens, W.A. *Comparison of Methods For Detecting the Impacts of Instructional/Test Misalignment*. AERA Annual Meeting, New Orleans, April 1988.
- Clarizio, H.F. & Phillips, S.E. *Comparison of Standard Score and Regression Methods of Identifying Learning Disabled Students*. AERA Annual Meeting, New Orleans, April 1988.
- Mehrens, W.A., McLarty, J.R., Rakow, E., & Phillips, S.E. *Conjunctive and Compensatory Models for Career Ladder Decisions: An Empirical Investigation*. NCME Annual Meeting, New Orleans, April 1988.
- Phillips, S.E. & Mehrens, W.A. *Relating Achievement Test Scores and Item Statistics to Instructional Validity*. NCME Annual Meeting, Washington, D.C., April 1987.
- Mehrens, W.A. & Phillips, S.E. *Conjunctive Versus Compensatory Models For Teacher Licensure Decisions: A Monte Carlo and Logical Investigation*. AERA Annual Meeting, Washington, D.C., April 1987.
- Lehmann, I.J. & Phillips, S.E. *Teacher Competency Examination Programs: A National Survey Revisited*. NCME Annual Meeting, Washington, D.C., April 1987.
- Phillips, S.E. & Mehrens, W.A. *The Effects of Curricular Differences on the Achievement Test Scores of Special Groups*. AERA Annual Meeting, San Francisco, April 1986.
- Phillips, S.E. & Clarizio, H.F. *Some Limitations of Standard Scores in Diagnosing Learning Disabilities*. AERA Annual Meeting, San Francisco, April 1986.

- Mehrens, W.A. & Phillips, S.E. *Sensitivity of Special Group Item Statistics to Curricular Validity*. AERA Annual Meeting, San Francisco, April 1986.
- Phillips, S.E. & Mehrens, W.A. *Achievement Test Curricular Multi-dimensionality at the Item and Objective Level*. AERA Annual Meeting, Chicago, April 1985.
- Mehrens, W.A. & Phillips, S.E. *Sensitivity of Item Statistics to Instructional Validity*. NCME Annual Meeting, Chicago, April 1985.
- Lehmann, I.J. & Phillips, S.E. *Teacher Competency Examination Programs: A National Survey*. NCME Annual Meeting, Chicago, April 1985.
- Clarizio, H.F. & Phillips, S.E. *Sex Bias in Diagnosing Learning Disabled Students*. AERA Annual Meeting, Chicago, April 1985. (Presented by Phillips).
- Phillips, S.E. *Quantifying Errors in IRT Equating Methods*. AERA Annual Meeting, New Orleans, April 1984.
- Phillips, S.E. *Fixed Versus Estimated Lower Asymptotes in the Three-Parameter IRT Model*. NCME Annual Meeting, New Orleans, April 1984.
- Phillips, S.E. & Anderson, A.E. *Comparison of the Parameter Recovery of the New and Old Versions of LOGIST with Simulated Data*. NCME Annual Meeting, New Orleans, April 1984.
- Phillips, S.E. & Mehrens, W.A. *Detecting Curricular Multidimensionality in Achievement Test Data*. AERA Annual Meeting, New Orleans, April 1984.
- Phillips, S.E. *Logistic Achievement Test Scaling and Equating with Fixed Versus Estimated Lower Asymptotes*. NCME Annual Meeting, Montreal, April 1983.
- Phillips, S.E. & Anderson, A.E. *LOGTRUE: A Computer Program For Test Equating with Item Response Theory*. AERA Annual Meeting, Montreal, April 1983.
- Phillips, S.E. *Comparison of Latent Trait and Traditional Methods in the Equating Phase of a Scaling Operation on Achievement Tests*. Psychometric Society Meeting, Montreal, May 1982.
- Phillips, S.E. *The Metrics are Here*. A series of 26 three-minute metric education programs shown on the KCRG TV Morning Show, Cedar Rapids, Iowa, 1982.
- Phillips, S.E. *Development of a Model For Adult Metric Education*. National Adult Basic Education Conference, Hollywood Beach, Florida, 1978.
- Phillips, S.E. *Overview of Project Metric*. National Metric Education Conference, Providence, Rhode Island, 1978.

Speeches

- Phillips, S.E. (1993). *Overview of Legal Implications of High-Stakes Performance Assessment*, Littleton, CO public schools, November 1993.
- Phillips, S.E. (1991). *Legal Issues and Educational Assessment*. Policy Seminar for State Legislators in Michigan, Lansing, October 17, 1991.
- Phillips, S.E. (1990). *Legal Issues in Testing*. Workshop for The Psychological Corporation, San Antonio, TX.
- Phillips, S.E. (1990). *The Rasch Model For Test Development and Equating*. Workshop for The Psychological Corporation, San Antonio, TX.
- Phillips, S.E. & Fremer, J. (1990). *The Revised Scholastic Aptitude Test*. WKAR Public Radio Station, East Lansing, MI.
- Phillips, S.E. (1987). *A Rasch Model Adaptation for Essay Calibration*. University of Tokyo, Tokyo, Japan.
- Phillips, S.E. (1987). *Simulation of Conjunctive Versus Compensatory Models for Teacher Competency Testing*. Presentation at Niigata University, Niigata, Japan.
- Phillips, S.E. (1987). *Equating with the Rasch Model*. Presentation to the New Jersey State Department of Education Technical Advisory Committee Meeting, Princeton, NJ.
- Phillips, S.E. (1987). *Standardized Achievement Tests & Curricular Differences*. Presentation at Arizona State University, Phoenix, AZ.
- Phillips, S.E. (1987). *Research on the Effects of Curricular Differences on Standardized Achievement Test Scores*. Presentation to the Marketing Department of The Psychological Corporation, San Antonio, TX.
- Phillips, S.E. (1987). *Item Response Theory Equating*. Presentation to the Texas State Department of Education, Austin, TX.
- Phillips, S.E. (1987). *The Curricular Validity of Achievement Tests*. Presentation at the University of Arizona, Tuscon, AZ.
- Phillips, S.E. (1987). *Developing Criterion-Referenced Tests*. Presentation to the Albuquerque (NM) Public Schools testing staff.
- Phillips, S.E. (1987). *Curricular Differences and Standardized Achievement Tests*. Presentation at the University of Virginia, Charlottesville, VA.

Phillips, S.E. (1987). *Linking Tests with the Rasch Model*. Presentation to the New Jersey State Department of Education, Trenton, NJ.

Phillips, S.E. (1987). *Developing a Test Blueprint*. Presentation to the Lansing Evaluation Services and Mathematics Staff, Lansing, MI.

Phillips, S.E. (1986). *Standardized Tests and Different Curricula*. Presentation to education faculty and students at Niigata University, Japan.

Phillips, S.E. (1986). *The Rasch Model*. Presentation to faculty in educational psychology at the Institute for Teacher Training and Ministry of Education test development staff, Republic of Singapore.

Phillips, S.E. (1986). *Using the Rasch model for the MEAP Tests*. Meeting with Michigan Department of Education assessment staff, Lansing, MI.

PROFESSIONAL SERVICE

- Admitted to Michigan Bar & Federal Bar for Eastern District of MI
- ETS Visiting Committee, 1995-97 (Audits selected testing programs for adherence to ETS and professional standards.)
- NCME Newsletter Board & Legal Corner Column
- 1992-95 Author's Committee, Education Law Reporter
- NOLPE 1992 Annual Convention Program Committee
- NCME Legislative Action Committee
- Co-chair of 1986 NCME Annual Meeting Program
- Consultant to the Arkansas, Alabama, Michigan, Minnesota, Mississippi, New Jersey, North Carolina, and Oregon Departments of Education, Texas Education Agency, Kentucky Office of Educational Accountability, Michigan Law Enforcement Officer Training Council, Minnesota Bar Examiners, New York Bar Examiners and GED Testing Service on student assessment, teacher certification, professional licensure, testing accommodations, assessment legislation, and/or legal challenges; Lansing, Weld County & Littleton, CO Public Schools on student performance assessment; and American Association of Medical Colleges on the MCAT writing assessment
- Reviewer for APPLIED MEASUREMENT IN EDUCATION, APPLIED PSYCHOLOGICAL MEASUREMENT, EDUCATIONAL RESEARCHER, JOURNAL OF EDUCATIONAL MEASUREMENT, JOURNAL OF EDUCATIONAL STATISTICS, PSYCHOLOGICAL BULLETIN, EDUCATIONAL MEASUREMENT: ISSUES & PRACTICE, EDUCATION LAW REPORTER

- Reviewer of Grant Applications for U.S. Office of Education
- Reviewer of measurement/evaluation textbooks for Longmann & Merrill
- Chair of Hearing Board for student grievance against a department faculty member.
Convened meetings of the board, conducted a hearing, and wrote a final opinion on the matter which contributed to dismissal of related charges being investigated by a national professional and accreditation organization.
- MQM Coordinator 1995-96 academic year.
Included convening meetings, course scheduling, student admissions, coordinating the MQM Internship Program, student evaluations, coordinating measurement comprehensive examinations and administrative paperwork for the program.

AWARDS

Graduated from law school summa cum laude with the James E. Burns Memorial Award for Scholarship Excellence

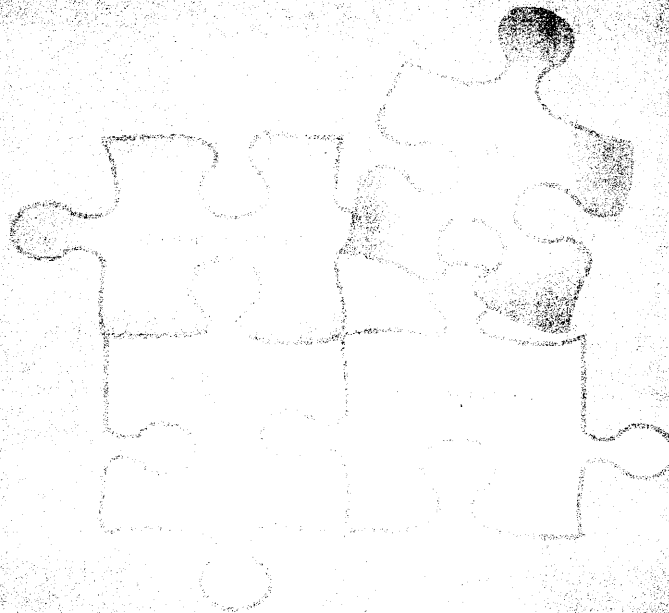
Edward G. Rakow law scholarship awarded annually to the outstanding student from each of the five law schools in Michigan

Alternate for National Academy of Sciences Spencer Fellowship, May 1986

Runnerup for AERA Outstanding Dissertation Award, April 1983.

Academic Accountability:

WILLIAMS V. STATE OF CA
EXPERT WITNESS REPORT



S. E. PHILLIPS, CONSULTANT

APRIL 2003

EXECUTIVE SUMMARY

The purpose of this report is to provide an overview of the California academic accountability system, to document how that system is improving over time, and to respond to the Russell Report.

The Accountability System

State accountability programs have three major components: content standards, assessment instruments, and consequences. California standards have received high marks in state comparisons by the Fordham Foundation and *Education Week*. While tests designed to measure California's challenging content standards were under development, the state administered the valid and reliable, nationally-normed, *Stanford Achievement Test*. The 1999 *Accountability Act* created an Academic Performance Index (API) to evaluate schools for awards and interventions.

Policy decisions for the accountability program are made by the appointed *Board* and implemented by the *Department* administered by the elected *Superintendent*. The *Accountability Act* provides for a representative and diverse *Advisory Committee* to counsel the *Superintendent* and the *Board*. A Technical Design Group was established by the *Committee* to develop calculation rules for the API.

The API is a summary, school-level measure of student academic performance in the content areas of *language arts*, mathematics, science and history/social science. The API provides scores for schools on a scale ranging from 200 to 1000, ranks schools on a scale ranging from 1 to 10, sets a statewide interim performance target of 800, assigns schools specific growth targets for future improvement, and provides comparisons between schools with similar characteristics.

To be eligible for awards, a school must meet an overall API growth target, meet comparable improvement growth targets for each numerically sig-

The API, created by the Accountability Act, is a summary, school-level measure of student academic performance in 4 content areas.

Schools meeting specific API criteria are eligible for awards; below-average schools that do not meet their growth targets are eligible for the intervention program.

nificant ethnic and socioeconomically disadvantaged subgroup, have a participation rate of at least 90-95% overall and 85% per content area, have less than 10% parent exemptions, and have *no* reports of adult testing irregularities. Initially, three different programs provided monetary awards; only non-monetary awards have been given since 2002.

Schools scoring in the bottom half of the API distribution and *not* meeting their growth targets are eligible for the intervention program. Participating schools are funded to implement an action plan developed by a school site team and an external evaluator. A participating school that has not demonstrated significant growth after 3 years is identified as low-performing and the *Superintendent* is authorized to intervene.

The Evolving and Improving API

Since its inception in 1999, the API has evolved from including only the *Stanford Test* to also including California Standards Tests in *Language Arts*, *Mathematics*, and *History*, and the *High School Exit Exam*. In that process, the weight of the *Stanford Test* in the API decreased from 100% to 20% at the elementary and middle school levels and from 100% to 12% at the high school level. Other measures will be added when they are judged valid and reliable as required by the *Accountability Act*.

The Russell Report

The Russell Report examines the California testing and accountability systems, including the API. It critiques the system with respect to choice of indicators, validity of test data, policy decisions, error in the system, consequential validity, lessons from other states, public opinion and teacher variables.

Unfortunately, the Russell Report is heavy on conjecture, short on supporting evidence and dismissive of contradictory information. Contrary to its assertions, data indicate alternative API policies were debated, students have benefited, and Rhode Island is not an appropriate model for California.

*The API is a
work in
progress.*

*Unfortunately,
the Russell
Report is heavy
on conjecture,
short on
supporting
evidence and
dismissive of
contradictory
information.*

TABLE OF CONTENTS

Introduction	1
Summary of Qualifications	1
California Consulting	2
Overview of the CA Academic Accountability Program	3
Content Standards	3
Assessment Instruments	4
<i>Assessment Act Provisions</i>	4
<i>The Stanford Test</i>	5
<i>Validity</i>	6
<i>Content Validity Evidence</i>	6
Augmentation of the <i>Stanford Test</i>	7
The California Standards Tests	9
The <i>California Survey Test</i>	9
The <i>High School Exit Exam</i>	10
Summary	10
Consequences	11
<i>Accountability Act Provisions</i>	11
The PSAA Advisory Committee	13
Calculating an API	14
<i>API Calculation Steps</i>	15
<i>Subgroup APIs</i>	15
Ranking API Performance	16
Calculating Growth Targets	16
<i>Overall School Growth Targets</i>	16
<i>Comparable Improvement Growth Targets</i>	17
Calculating Annual Growth	17
Summary of the API	20
API Awards	20
Interventions	22
Evaluation Reports	23
Alternative Accountability System	24
Evolution of the API	25
Addition of Standards Tests	25
Future Plans	26
Response to Russell Report	28
Choice of Indicators	28
<i>Stated Intent of the Accountability Act</i>	29
<i>Improved Ethnic and SES_D Subgroup Achievement</i>	29
Statewide Data	30
Data for Individual Districts & Schools	31
<i>Role of Inputs in the Current Accountability System</i>	32
Disclosure of Information	32
Intervention Process Uses Inputs	34
<i>Evidence From an Unsuccessful, Court-Imposed Inputs Experiment</i> ..	35
Validity of Test Data	37
<i>Testing CA Content Standards</i>	38

<i>Replacing the Stanford Test with the California Test</i>	39
<i>Plans for Additional Non-Test Indicators</i>	40
<i>NAEP Comparisons</i>	41
Policy Decisions	42
<i>General Policy Disagreements</i>	43
<i>Alleged Unfairness of API Policies</i>	45
<i>Second-Guessing Policymakers</i>	46
Reasonable, Judgmental Process	47
Calculation Rationales	47
Balancing Policy Tradeoffs	49
Reconsideration of API Policies	49
Russell Alternatives Considered and Rejected	50
Summary	50
<i>Consistency with Federal Law</i>	51
<i>Aggregation Across Grades</i>	52
Error in the Accountability System	53
<i>Test Score Error</i>	53
<i>API Error</i>	54
<i>Reporting Errors</i>	56
Consequential Validity	56
<i>Limitations of Teacher Survey</i>	58
<i>Potential Positive Consequences</i>	59
Teaching the Content Standards	59
Efficient and Targeted Use of Available Resources	61
Improvement of Low-Performing Subgroups	61
Better Decision-Making	62
<i>Attenuating Alleged Negative Consequences</i>	62
Detecting & Deterring Unethical Behavior	62
Allocating Responsibility for Explaining School Performance ..	65
Recognizing Lack of Causation in Dropout and Retention Rates	66
Following Legal Requirements for Special Education Classification	67
<i>Consequential Validity Summary</i>	68
Lessons From Other States	69
Public Opinion and External Evaluations	71
<i>California Opinion Data</i>	72
Teacher Variables	72
<i>Correlation with SES</i>	73
<i>Data for Schools Attended by Named Plaintiffs</i>	75
Conclusion	76

LIST OF CHARTS

- | | |
|-----------|--|
| Chart 1 | Overview of the CA Academic Accountability System |
| Chart 2 | Organizational Chart for the CA Academic Accountability System |
| Chart 3 | Timeline of <i>Board</i> Actions Involving the API |
| Chart 4 | API Policy Reports Prepared for the <i>Board</i> by the <i>Advisory Committee & Tech Group</i> |
| Chart 5 | Sample 2000 API Elementary School Calculation |
| Chart 6 | API Reporting Cycles |
| Chart 7 | Purpose of the Scaling Factor in the API |
| Chart 8 | Sample 2002 Base API Calculation for a High School |
| Chart 9 | API Six-Year Plan |
| Chart 10 | Summary of API Award Programs |
| Chart 11a | School Report Card Summary for Coronado Elementary School |
| Chart 11b | 2002 API Growth Report for Coronado Elementary School |
| Chart 12a | Sample <i>Stanford Test</i> & Standards Tests Student Report |
| Chart 12b | Sample <i>Exit Exam</i> & <i>Stanford Test</i> School Reports |
| Chart 12c | 1998-2001 <i>Intervention Program</i> Schools' APIs & Test Scores |
| Chart 13 | Relationship Between API Improvement, School Size & Standard Error |
| Chart 14a | 2002-1998 NPR Gains in "Average" <i>Stanford Test</i> Scores |
| Chart 14b | 2001-1999 Gains in Average Percent Correct for Standards Tests |
| Chart 15 | 2002 State Base <i>Stanford Test</i> Only versus Standards Tests Only APIs |
| Chart 16a | 2002-1998 <i>Stanford Test</i> Reading Growth by District for Selected Grades |
| Chart 16b | 2002-1998 <i>Stanford Test</i> Mathematics Growth by District for Selected Grades |
| Chart 17 | CA Subgroup & US NAEP Changes in Percent Proficient & Above |
| Chart 18 | 2000-2002 API Awards Eligibility by Decile |
| Chart 19 | Awards & Intervention Funding for Schools Attended by Named Plaintiffs |
| Chart 20 | Percents of Credentialed Teachers by Type for Schools Attended by Named Plaintiffs |

LIST OF TABLES

Table 1	Match of <i>Stanford Test & California Test</i> to CA Content Standards
Table 2a	Statewide <i>Stanford Test</i> Summary Data
Table 2b	Statewide Standards Test Summary Data
Table 2c	2002 Statewide <i>Stanford Test</i> Only versus Standards Tests Only APIs
Table 3a	Data for Individual School Districts & Statewide Data
Table 3b	Data for Individual Schools Attended by Named Plaintiffs
Table 4	Summary Status of <i>Intervention Program</i> Cohort I Schools
Table 5	Comparison of California NAEP & Standards Tests Trend Data
Table 6a	Summary of Statewide API School Performance
Table 6b	API Awards Eligibility by Deciles & Reasons for <i>NO</i> 2002 School Growth API
Table 7	Ratings of Selected State Standards & Accountability Systems
Table 8	Fordham Evaluation of State Standards & Accountability Systems
Table 9	Demographic Data for Selected States
Table 10	Test Score Data for Selected States
Table 11	California Public Opinion Poll Data
Table 12	2001 Mean Number of Teachers with Full & Emergency Credentials & SES ₀ API Gains for Elementary Schools (Grades 2-6) by Decile
Table 13a	Summary API Data for CA Elementary Schools Attended by Named Plaintiffs
Table 13b	Summary API Data for CA Middle Schools Attended by Named Plaintiffs
Table 13c	Summary API Data for CA High Schools Attended by Named Plaintiffs
Table 14	Summary Demographic Data for CA Schools Attended by Named Plaintiffs
Table 15	Award Probabilities & False Positives

List of Acronyms

acronym	meaning	short form used in this report
API	Academic Performance Index	<i>API</i>
CAHSEE	California High School Exit Examination	<i>Exit Exam</i>
CAPA	CA Alternate Performance Assessment	<i>Alternate Assessment</i>
CAT6	California Achievement Test Sixth Edition	<i>California Test</i>
CDE	California Department of Education	<i>Department</i>
CEC	California Education Code	<i>Education Code</i>
CELDT	CA English Language Development Test	<i>English Test</i>
CSIS	California Student Information System	<i>Information System</i>
CSR	Code of State Regulations	<i>Regulations</i>
ELA	English language arts/reading	<i>Language Arts</i>
GPA	Governor's Performance Award Program	<i>Award Program</i>
II/USP	Immediate Intervention/Underperforming Schools Program	<i>Intervention Program</i>
NAEP	National Assessment of Educational Progress	<i>National Assessment</i>
NCLB	No Child Left Behind Act	<i>Federal Law</i>
PSAA	Public School Accountability Act	<i>Accountability Act</i>
SAT9	Stanford Achievement Test Ninth Edition	<i>Stanford Test</i>
SBE	State Board of Education	<i>Board</i>
S_D	Socio-economically Disadvantaged	<i>SES_D</i>
SPI	Superintendent of Public Instruction	<i>Superintendent</i>
STAR	Standardized Testing and Reporting	<i>Assessment Act</i>

**ACADEMIC ACCOUNTABILITY:
WILLIAMS V. STATE OF CA EXPERT WITNESS REPORT**

INTRODUCTION

The purpose of this report is to provide an overview of the California academic accountability system and to document how that system is evolving and improving over time. This report also responds to the report by plaintiffs' expert, Michael Russell.

In preparing this report, I have reviewed plaintiffs' statement of liability; reports by plaintiffs' experts; the Public Schools Accountability Act (PSAA) and the Student Testing and Reporting (STAR) Act; State Board of Education (SBE) minutes and agenda items; PSAA Advisory Committee minutes; informational reports about the California accountability program, assessment system, academic performance index (API), school test and API results, and other information available on the California Department of Education (CDE) website; and research and evaluation reports related to California standards, assessments and the API. I also interviewed the director of the CDE Policy and Evaluation Division, and a Co-chair of the Technical Design Group and member of the PSAA Advisory Committee.

The professional opinions set forth in this report are based on review of the information listed above, familiarity with the California student assessment and school accountability programs, and my professional knowledge obtained from extensive training, experience and scholarship

in the areas of psychometrics and assessment law. A summary of my professional qualifications follows.

SUMMARY OF QUALIFICATIONS

Since 2000, I have been an independent, educational consultant specializing in psychometrics and assessment law. Formerly, I was a professor of educational measurement at Michigan State University for 18 years. My educational training includes a PhD in educational measurement and statistics from the University of Iowa and a JD degree from Thomas M. Cooley Law School in Lansing, Michigan.

As a member of the graduate faculty at Michigan State University, I taught courses in psychometrics and statistics with a specialization in legal and policy issues. In addition, I taught an elective on Legal Aspects of Educational Assessment at Cooley Law School. Prior to joining the Michigan State University graduate faculty, I worked in the test division of Riverside Publishing Company and for the American College Testing Program.

Recently I served as an expert witness and consultant in the Texas *GI Forum* lawsuit, in which the state of Texas successfully defended the exit level test required for high school graduation.¹ Other states where I have served as an expert witness

¹ *GI Forum v. Texas Education Agency*, 87 F. Supp. 667 (W.D. Tex. 2000).

include Alabama, California, Connecticut, Pennsylvania and Virginia, in cases involving assessment accommodations, assessment of English language learners, test security and graduation testing.² I have also served on Technical Advisory Committees or as an assessment law consultant for many statewide testing programs, including Arizona, California, Delaware, Florida, Hawaii, Kentucky, Massachusetts, Michigan, Minnesota, Mississippi, New Mexico, New York, Oregon, Nevada, New Jersey, Texas and Washington. My vita lists more than 35 publications and 60 professional presentations on assessment topics, including a handbook on legal issues in assessment designed to provide measurement specialists, administrators, legislators, policymakers, and others involved in assessment enterprises with a concise summary of the legal and policy implications of high-stakes assessment decisions.

In addition, I have made presentations on legal aspects of assessment issues to a variety of national forums including the NAGB/NCES Joint Conference on Standard Setting in Washington, DC, the CCSSO Large-scale Assessment Conferences, the AERA/NCME annual meetings, the National Organization on Legal Prob-

² Golden v. Birmingham Bd. of Educ. & Alabama Dept. of Educ., IDEA due process hearing, January 1994; SFUSD et al. v. State Bd. of Educ. et al., Case No. 99409 (Sup.Ct. Cal. Dec. 2000); Chapman et al. v. Calif. Dept. of Educ. et al, Superior Court for the State of Calif., County of Alameda, Case No. 2002049636 (pending); Chapman et al. v. Calif. Dept. of Educ. et al, U.S. District Court, Northern District of Calif., Case No. C01 1780 CRB (pending); Fairfield School District Employment Action (1996); Doe v. NBME, U.S. District Court, Eastern District of Pennsylvania, Case No. 99-4532 (pending); Pandazides v. Virginia Bd. of Educ., 752 F. Supp. 696 (E.D. Va. 1990), *rev'd*, 946 F.2d 345 (4th Cir. 1991).

lems in Education annual meeting, the National Conference of Bar Examiners and several regional conferences including the Michigan School Testing Conference, the Colorado, Texas and Mississippi Assessment Conferences and the Wisconsin Outcomes Based Education Conference. I have also served on the ETS Visiting Committee, the Author's Committee for the Education Law Reporter and the editorial board of the NCME newsletter, including contribution of a number of columns on legal issues in assessment. Prior to publication of the revised *Test Standards* in 1999, I was asked by NCME to review and comment on the Chapter on *Testing Individuals with Disabilities*.³

California Consulting

I have served as a consultant to the California State Board of Education (*Board*) on the selection of a standardized test pursuant to California Education Code (CEC) §§ 60600-60647 and to the California Department of Education (*Department*) on the high school graduation test. My work as a consultant to the *Board* occurred in November of 1997 and involved a psychometric evaluation of proposals submitted by test publishers for the standardized testing component of the Standardized Testing and Reporting Program. My work with the *Department* includes a presentation on Setting Performance Standards (March 1998), attending Advisory Committee meetings (November 1998; January 1999), a presentation on Opportunity to Learn

³ See *Test Standards*, *infra* note 13.

and Testing Accommodations (November 1999), and continuing consultation on the high school graduation test beginning in September 2000. My most recent role as a consultant to the *Department* and member of the *Expert Panel on Assessment* has involved providing technical expertise on a variety of assessment issues.

OVERVIEW OF THE CA ACADEMIC ACCOUNTABILITY PROGRAM

Academic accountability programs at the state level have three major components. They include: (1) academic content standards by grade level in core subjects (e.g., reading, mathematics); (2) assessment instruments to measure achievement of those standards; and (3) consequences (rewards and interventions) for successful and unsuccessful schools.

In California, academic content standards have been adopted for English language arts/reading (ELA or *language arts*), mathematics, science and social science. The assessment system used to measure achievement of the California content standards consists of the Standardized Testing and Reporting (STAR) program and the California High School Exit Examination. The STAR program includes a nationally-normed, standardized achievement test and California standards tests in *language arts* and mathematics. Additional California standards tests in science and social science are currently administered at the high school level and are under develop-

ment for earlier grades. The consequences component of the accountability program, consisting of an academic performance index (API) and associated rewards and interventions, was created by the Public Schools Accountability Act (PSAA or *Accountability Act*).

The California Department of Education (CDE or *Department*), under the supervision of the elected Superintendent of Public Instruction (SPI or *Superintendent*), is responsible for legislative implementation and administration of the accountability program. Policy decisions for the accountability program are made by the State Board of Education (SBE or *Board*), whose members are appointed by the governor. Additional leadership and coordination is provided by the governor's Secretary of Education.

Chart 1 presents an overview of the major components of the California academic accountability system. Standards, assessments and consequences of the California academic accountability system as prescribed in the *Accountability Act* and other legislation are considered in more detail in the sections that follow.

CONTENT STANDARDS

California's accountability program for holding its schools responsible for student achievement of important academic skills in core subjects began in the mid 1990s with the formation of stakeholder committees to develop content standards by grade level and subject matter that detailed

the knowledge and skills that all students should achieve. Members of the stakeholder committees who provided input and debated options included educators, business leaders, parents and the general public representing all regions of California. This effort culminated in the adoption of Reading/English Language Arts (*language arts*) and Mathematics Content Standards in 1997/98 and Science and History/Social Science Content Standards in 1999.⁴

The California content standards have received high marks in state comparisons by the Fordham Foundation and *Education Week*. In 2000, Fordham graded California's *language arts*, mathematics, history and science content standards an "A" and awarded a "C" for geography standards (included in history/social science).⁵ Overall, California's standards rated an "A-" from Fordham, the top state grade and only "A" given.⁶

In 2003, the California standards and accountability system received a grade of "B+" from *Education Week*.⁷ Although the Fordham criteria for judging state content standards were detailed and specific, the *Education Week* criteria were more general and primarily evaluated the presence or absence of certain features such as test item types.⁸

⁴ SBE Minutes & Agenda Items, ELA #22, Nov. 14, 1997; Math #19, Dec. 11, 1997; Science & History/Social Science, #34, Oct. 9, 1998.

⁵ Finn, C. & Petrilli, M. (Ed.), *The State of State Standards 2000*, The Thomas B. Fordham Foundation, January 2000, p. 3. See also Table 1 *infra*.

⁶ *Id.*

⁷ *Quality Counts*, *Education Week*, XXII (17), January 9, 2003, p. 84-85, 102.

⁸ *Id.* at 102; See footnotes in Table 1, *infra*.

In its review, Fordham ranked California's *language arts*, history, mathematics, and science content standards first in the nation stating:

[California] has clear, specific, and measurable standards, and addresses all areas of [*language arts*] well and comprehensively. ... The *California History Standards* exemplify "best in nation" for history standards writing, presentation, and content. ... California now boasts one of the best science standards presently available.⁹

ASSESSMENT INSTRUMENTS

Assessment Act Provisions

The Leroy Greene California Assessment of Academic Achievement Act established the Standardized Testing and Reporting (STAR or *Assessment Act*) program and its guidelines. Relevant sections state:

It is the intent of the Legislature in enacting this chapter to provide a system of individual assessment of pupils that has, as its *primary purpose*, assisting pupils, their parents, and teachers to *identify individual academic strengths and weaknesses, in order to improve teaching and learning*. It is further the *intent* of the Legislature in enacting this chapter to *determine the effectiveness of school districts and schools, as measured by the extent to which pupils demonstrate knowledge of the fundamental academic skills, as well as the ability to apply those skills*.

...

(a) There is hereby established the [*Assessment Act*] program.

⁹ Fordham Report, *supra* note 5 at 8, 12, 20, 23, 34.

(b) Commencing in the 1997-98 fiscal year and each fiscal year thereafter, ... *each school district*, ... shall administer to *each of its pupils* in grades 2 to 11, inclusive, before May 15, *the achievement test designated by the [Board]*.

(c) [Makeup days].

(d) [Testing in grades 1 and 12].

(e) *Individuals with exceptional needs* who have an *explicit provision* in their individualized education program that exempts them from the testing requirement ... *shall be so exempt*.

(f) At the school district's option, pupils of limited English proficiency who are enrolled in any of grades 2 to 11, inclusive, may take a second achievement test in their primary language. ...

(g) Pupils of limited English proficiency who are enrolled in any of grades 2 to 11, inclusive, shall be required to take a test in their primary language if such a test is available, if fewer than 12 months have elapsed after their initial enrollment in any public school in the state.

...

Based upon a review of the achievement tests submitted and the recommendation made by [the *Superintendent*], [the *Board*], in its sole discretion, based on the [psychometric, feasibility, cost and experience criteria stated in § 60644], shall designate for use as part of the [Assessment Act] Program a *single test* in grades 2 to 11, inclusive, no later than November 14, 1997.

...

In *designating an achievement test*, [the *Board*] shall adopt *only a nationally normed test* and shall consider each of the following criteria:

(a) Ability of the publisher to produce *valid, reliable* individual pupil scores.

(b) Quality and age of empirical data supporting national norm referenced data analysis of the proposed assessment. ...

(c) Ability to report [individual student scores, aggregated test results, and disaggregated scores for ethnic subgroups and English Language Learners ...] ...¹⁰

Title 5 of the California Code of Regulations adopted by the *Board* provides specific rules for the administration of the *Assessment Act* Program designated achievement test. For the first five years of the program, the Stanford Achievement Test Ninth Edition (SAT9 or *Stanford Test*) was the achievement test adopted by the *Board*. Beginning in spring 2003, the designated norm-referenced achievement test will be the survey version of the California Achievement Test Sixth Edition (CAT6 or *California Test*).

The Stanford Test

The *Stanford Test* is a nationally-normed, standardized achievement test that has been administered annually to all students in grades 2 through 11 in California. The test measures academic skills in reading, mathematics, language, and spelling in elementary and middle grades, and in reading, language, mathematics, science and social science at the high school level. These subtests are based on knowledge and skills commonly included in the grade level curricula of a majority of school districts in the United States.¹¹

¹⁰ CEC § 60602(a); § 60640(a)-(g); § 60642(b); § 60644(a)-(c); emphasis added.

¹¹ Districts may elect to administer optional *Stanford* subtests at other grade levels.

In reading, the tested skills include vocabulary and comprehension; in math, they include problem solving and procedures. The test items are presented in multiple-choice format and often include pictures, graphics, or other stimulus materials. Except for elementary students in grades 2 and 3 who mark their answers in a machine-scorable test booklet, students mark their answers on a separate answer sheet. All test directions, questions and stimulus materials are written in English.¹²

Validity

Three national professional organizations have collaborated to produce consensus standards for educational and psychological testing.¹³ The *Test Standards* state:

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. . . . It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself.¹⁴

Validity refers to the weight of accumulated evidence supporting a particular use of test scores. The Stanford Test (and California Test) scores are used by the state to determine whether schools are meeting their growth targets for academic im-

¹² See Stanford Achievement Test Ninth Edition Technical Manual.

¹³ American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), *Standards for Educational and Psychological Testing*. Washington, DC (1999) [hereinafter referred to as *Test Standards*].

¹⁴ *Id.* at 9.

provement and by schools and parents to identify individual students' strengths and weaknesses. The most important evidence of validity in this situation is a measure of the degree to which the items on each content area test measure the knowledge and skills identified by California as important for all students to achieve. This type of validity evidence is referred to as *content validity* evidence.

Content Validity Evidence

Standards 1.6, 1.7, 3.2-3.3, 3.5-3.9, 3.11 and 13.3 from the 1999 *Test Standards* deal specifically with issues related to content validity evidence. These Standards require that the purpose of the test, procedures used to specify the content domain, the qualifications of content experts, and the procedures used to obtain expert judgments be clearly documented. These requirements for developing content validity evidence are described more fully below.

As indicated in the *Test Standards*, content validity evidence for an achievement test is *typically obtained by professional judgment*. Based on the purpose of the test, a diverse panel of content experts is asked to identify an age-appropriate, testable domain of academic subject matter from the state content standards, to develop a set of test specifications which identifies the specific knowledge and skills to be sampled from the domain, and to specify the proportional weight to be given to each sampled content area. This information constitutes the *Test Specifications*.

After trained item writers have produced a set of items based on the test specifications, the diverse panel of content experts is asked to review each potential test item and determine whether it measures the intended subject matter skill. As part of this review, these content experts also check the correctness of the keyed answer, check for ambiguities in wording and other potential item flaws, evaluate the appropriateness of the content and difficulty of the item for the intended grade level, and identify any inappropriate or potentially offensive language or content that might impair accurate assessment of ethnic minority and socio-economically disadvantaged (SES_D) students.

The edited items are then field tested on a sample of students, item statistics are calculated and the items are evaluated again. During this second review of items, content experts re-examine the match of the item to the skill it is supposed to measure in the context of item data from the field test, including consideration of differential performance by ethnic minority subgroups. Test forms are constructed based on the content specifications of the test and are then administered to representative national samples of students to develop the test norms.

The *Stanford Test* was developed using the extensive test development procedures described above. In addition, when the *Board* considered the publisher's proposal for adoption of the *Stanford Test*, the *Board* was furnished with documents that matched the *Stanford Test* content to

the California state standards in each subject area. The quality of the test development effort as documented by the *Stanford Test* publisher, the information on the match of the *Stanford Test* to the state content standards, and evaluations by the *Board's* independent evaluators provided the *Board* with the necessary information to judge the content validity of the *Stanford Test*.

The *Board's* independent evaluators rated each publisher's test proposed for the *Assessment Act* Program on a variety of factors. In particular, evaluators were asked to rate each proposal on each of the statutory criteria including the "ability to produce valid, reliable, individual pupil scores" and the "quality and age of empirical data supporting national norm-referenced data analysis of the proposed assessment."¹⁵ Evaluators were also asked to provide comments on strengths and weaknesses relative to each of the criteria and to provide additional comments to assist the *Board* in evaluating the proposals.

Based on extensive information provided by the publisher, the *Board's* independent evaluators (of which I was a member) judged the *Stanford Test* to be *valid for the assessment uses described in the STAR statute*. In November 1997, the *Board* adopted the *Stanford Test* for these purposes.¹⁶

¹⁵ CEC § 60644(a) & § 60644(b).

¹⁶ SBE, Minutes and Agenda Item #23, Nov. 14, 1997.

Augmentation of the *Stanford Test*

As indicated, the *Stanford Tests* in language arts and mathematics measure common national skills at each grade level. Many of these skills are consistent with the content skills specified in the California Content Standards. Many others represent prerequisite or enabling skills from previous grades.

However, because California chose to adopt content standards that are more demanding than the average expectations nationally at the upper grade levels, some skills included in the state standards are not tested by the *Stanford Test*. To provide a more targeted assessment of the specific content skills for each grade level, the state constructed its own standards-based tests as required by statute. But because such standards-based tests required extended time for development, the state adopted an interim procedure for estimating achievement of the California Content Standards.

The *Stanford Test* had to be administered intact under standard conditions in order for the norms to be valid and usable. However, in the interim while separate tests were being developed to measure state standards, the state wanted to estimate student achievement of the California content standards. To do so, the state developed an augmented test that consisted of *Stanford Test* items judged to match the standards plus a set of separately-administered, additional items measuring content standards not covered by the *Stanford Test*.

These augmented tests, combining scores from selected *Stanford Test* items and additional standards items, served as the California content standards tests until separate tests were introduced into the API in reading/language arts in 2000 and in mathematics in 2001. However, even after the California Standards Tests were introduced, the *Stanford Test* scores remained important for providing national comparisons and for measuring skill levels of students at the lower end of the achievement distribution.

In the early years after state adoption of the new content standards, schools were still in the process of adjusting their instructional programs to include all of the new state content standards. During this period, the *Stanford Test* and the augmented standards tests measured content that schools should have already been teaching plus new content that was in the process of being integrated into the instructional program. Measuring both existing and new curricula provided schools with information about progress in implementing the new state content standards and motivation to complete the process expeditiously.

This practice is consistent with sound measurement theory specifying that educational tests should measure what students are expected to learn and what the test administrator (in this case, the state) wants to evaluate. The introductory text to the Validity chapter in the 1999 *Test Standards* states: "In educational program evaluations, . . . tests may properly cover material that receives

little or no attention in the curriculum, as well as that toward which instruction is directed"¹⁷

The school API measures derived from *Stanford Test* and Standards Tests scores are an example of educational program evaluation referred to in the *1999 Test Standards*. Together, the *Stanford Test* and augmented tests provided the state with measures of enabling/lower level skills that schools should already have been teaching and standards-based skills consistent with the new California Content Standards.

Data on the match of the *Stanford Test* to the California Standards is presented in Table 1. These data indicate that across the grade levels, most of the content strands are measured by the *Stanford Test*. In addition, 40-55 *language arts* and 15 mathematics *Stanford Test* items at each grade level were included in the augmented tests designed specifically to measure California Standards.¹⁸

The California Standards Tests

The California Standards Tests (CSTs or *Standards Tests*) are specifically designed to measure the California content standards. *Standards Tests* have been developed for *language arts*, mathematics, history/social science and science. The *Standards Tests* were developed by adding enough additional items to

¹⁷ *Test Standards*, *supra* note 13 at 12.

¹⁸ Note that the *Stanford Test* items were administered and scored intact and then responses from selected *Stanford Test* items were combined with the separately-administered augmented items to create a composite California Standards Test score.

the separate augmented test item sets to render them valid and reliable as stand-alone instruments.

Currently, the *language arts* and mathematics Standards tests consist of multiple-choice items. The history/social science and science Standards Tests also consist of multiple-choice items. A writing Standards Test utilizing essay items is administered in grades 4 and 7 as part of the *language arts* Standards Test.¹⁹

The California Survey Test

Beginning in 2003, the California Achievement Test 6th Edition Survey Test (*California Test*) will be administered in grades 2 through 11 in California. The survey test is a shorter form of the complete battery test. Scores from the *California Test* will be linked to performance on the previously-administered *Stanford Test*.

Similar to the *Stanford Test*, the *California Test* is a norm-referenced achievement test consisting of multiple-choice items with responses recorded on machine-scorable answer sheets. Subtests include Reading/Language Arts, Mathematics, Science and Social Studies consisting of 60, 25, 25, and 25 items, respectively, at the upper grade levels and requiring about 3 hours to administer. Test construction procedures and technical characteristics are similar to those for the *Stanford Test*.²⁰

¹⁹ See www.cde.ca.gov for more information.

²⁰ See CTB McGraw-Hill, *California Achievement Tests Technical Report*, submitted for the CA STAR program, March 2002.

Data summarizing the match of the *California Test* to the California academic content standards are presented in Table 1. These data indicate that in the elementary and middle school grades, the *California Test* covers approximately a third of the standards with 70% to 100% of the test items in *language arts* and mathematics aligning with the California Standards.

The High School Exit Exam

In 1999, the legislature established the California High School Exit Examination (CAHSEE or *High School Exit Exam*) requirement in *language arts* and mathematics.²¹ Effective with the Class of 2004, students must pass the *High School Exit Exam* to be eligible for a high school diploma. With the passage of the *High School Exit Exam* statute, the legislature changed the requirements for a high school diploma from a rite of passage based on attendance and credits to a skills-based approach based on common, statewide examinations testing academic skills in core subjects, plus specified course requirements such as Algebra I.

The *High School Exit Exam* measures achievement of a designated subset of the California standards in *language arts* in grades 8-10 and mathematics in grades 6-7 plus algebra²² adopted by the State Board

²¹ CEC § 60850 et seq. (1999).

²² Note that although traditional algebra I content is now part of the California math content standards for grade 8, many students do not take an algebra course until high school.

of Education in 2000.²³ Content weighting for the *High School Exit Exam language arts* and mathematics tests (e.g., percent of items measuring statistics, geometry, algebra, etc. in math) is determined by the exit test specifications adopted by the Board.²⁴ The selected content standards for the *High School Exit Exam* are a subset of the middle school and high school *language arts* and math standards that represent minimum academic expectations for all high school graduates.

The *Accountability Act* legislation provides for the inclusion of high school exit exam results in the API "when available and found to be valid and reliable for this purpose."²⁵ *High School Exit Exam* results became part of the base API for high schools beginning in the 2002-03 school year.

Summary

The *Stanford Test*, its successor, the *California Test*, and the Standards Tests are achievement tests that measure students' acquisition of specified skills at a particular point in time. *The skills measured by these achievement tests are sensitive to instruction and students' proficiencies are expected to improve over time.* The *Assessment Act* and *Accountability Act* provisions require specific student information to be collected and reported but impose *no*

²³ SBE, Minutes & Agenda Item #21, Sept. 7, 2000. See www.cde.ca.gov/statetests/cahsee for more information.

²⁴ *Id.*

²⁵ CEC § 52052(b).

negative consequences on any individual students.

This part of the accountability system is designed to hold the educators responsible for teaching the state content standards to all students. Only the *High School Exit Exam* has consequences for individual students, and the affected high school students are provided multiple opportunities for remediation and retesting.

CONSEQUENCES

Accountability Act Provisions

In 1999, the legislature passed the Public School Accountability Act (PSAA or *Accountability Act*).²⁶ The *Accountability Act* includes three major components:

1. Academic Performance Index (API),
2. Immediate Intervention/ Underperforming Schools Program (II/USP or *Intervention Program*), and
3. Governor's Performance Award Program (GPA or *Awards Program*).

In enacting a new accountability system for California, the legislature stated:

(a) The purpose of the California public school system is to provide for the academic development of each pupil and prepare each pupil, to the extent of his or her ability, to become a life-long learner, equipped to live and

succeed within the economic and social complexities of the 21st century.

(b) It is in the interest of the people and the future of this state to ensure that each child in California receives a high quality education consistent with all statewide content and performance standards, as adopted by the State Board of Education, and with a meaningful assessment system and reporting program requirements.

(c) Recent assessments indicate that many pupils in California are not now, generally, progressing at a satisfactory rate to achieve a high quality education.

(d) To remedy this, the state is in need of an immediate and comprehensive accountability system to hold each of the state's public schools accountable for the academic progress and achievement of its pupils within the resources available to schools.²⁷

The legislature further found that to be "promising and effective," such an accountability system requires the involvement of parents, educators, administrators, and local community members engaged in constructive collaboration to improve student achievement. To this end, the legislature stated an intent to encourage:

teacher preparation and consistent ongoing professional development that serves to develop competency in content and pedagogy and that allows teachers to effectively involve themselves in promoting school accountability; and

local community involvement in providing support for education and identifying causes of pupil failure and designing programs for remediation.²⁸

²⁶ CEC § 52050 et seq. The summary in this section includes amendments made by SB1552 (Ch. 695 of 2000). Note: the Budget Act of 2001 changed the minimum growth expectation from 1 point to 5 points. SBE Minutes & Agenda Item #11, Oct. 10, 2001.

²⁷ CEC § 52050.5.

²⁸ *Id.*

The *Accountability Act* implemented the school evaluation requirement of the Assessment Act by establishing a state Academic Performance Index (API) and expected growth targets to be calculated for each California public and charter school with an enrollment of at least 100 students, a state intervention program for selected schools performing below the state average and failing to meet their API growth targets, and rewards for schools that meet or exceed their overall growth and comparable improvement targets.²⁹ Comparable improvement involves separate performance calculations for numerically significant ethnic and SES_D subgroups.³⁰

The framework for the API, adopted unanimously by the *Board* in July 1999, outlined guiding principles, design features and uses for the API.³¹ The 12 guiding principles for development stated that the API:

1. must be technically sound.
2. must emphasize student performance, not educational processes.
3. must strive to the greatest extent to measure *content, skills, and competencies that can be taught and learned in school and that reflect the state standards.*
4. must allow for fair comparisons.
5. should *include as many students as possible in each school and district.*
6. must measure school performance and growth as accurately as possible.

²⁹ CEC § 52052(a),(c),(e),(f); CEC §§ 52053 & 52056.

³⁰ CEC § 52052(a)(2); § 52057(a).

³¹ SBE, Minutes & Agenda Item #26, July 15, 1999.

7. should strive in the long-term to measure growth based on student-level longitudinal data.
8. should be flexible and its component indicators should be stable.
9. should be understandable, particularly to educators and parents.
10. is part of an overall accountability system that must include comprehensive information which incorporates contextual and background indicators beyond those required by law.
11. should minimize burden.
12. should *support* local accountability systems.³²

The framework noted that it was the role of the *Superintendent* and the *Board* to establish policy priorities to resolve any conflicts among these principles.³³

Based on authority granted by the *Accountability Act* and the guiding principles listed above, in November 1999, the *Board* approved the 1999 base year API calculation as recommended by the *Advisory Committee* relying on the expertise of the Technical Design Group.³⁴ In this action, the *Board* approved the *Stanford Test* as the sole component of the index, adopted an API scale of 200-1000, and set an interim API goal for all schools of 800 (roughly

³² PSAA Advisory Committee, *Framework for the Academic Performance Index*, July 1999, p. 2-4.

³³ *Id.* at 2.

³⁴ SBE, Minutes & Agenda Item #23, Nov. 9, 1999. By unanimous vote of the members present, the *Board* also deleted a recommendation to permanently exclude English language learners that was not endorsed by the *Superintendent* or the Governor and was viewed as contrary to the *Accountability Act*. (Like all other students, English language learners are excluded from the API for their first year in a school district.)

the equivalent of the top of the proficient level on the National Assessment of Educational Progress (NAEP).³⁵

In addition, the *Board* set the overall growth target for schools with base year APIs less than 800 at 5% of the difference between the school API and 800, or 1 point, whichever is greater, and for schools with APIs at or above 800, maintenance of an API score of at least 800.³⁶ Comparable improvement was defined by the *Board* as growth for every numerically significant subgroup of at least 80% of the overall school growth target.³⁷

In specifying the students to be included in the API, the *Accountability Act* further provided that:

Only the test scores of pupils who were enrolled in a school district in the prior fiscal year may be included in the test results reported in the API.³⁸ Results of the achievement test and other tests [] shall constitute at least 60% of the value of the [API]. Before including high school graduation rates and [student and teacher] attendance rates in the [API, the Superintendent] shall determine the extent to which the data is currently reported to the state and the accuracy of the data.

...

Pupil scores from the following tests, when available and when found to be valid and reliable for this purpose,

³⁵ SBE, Minutes & Agenda Item #23, Nov. 9, 1999; CDE, The 1999 Base Year API, www.cde.ca.gov, p. 6-7.14.

³⁶ Later codified by SB 1552 (Ch. 695 of 2000) as CEC § 52052(c).

³⁷ SBE, Minutes & Agenda Item #23, Nov. 9, 1999; The 1999 Base Year API, *supra* note 35.

³⁸ Amended by CEC § 52052.3 to also include first year high school students from elementary schools that normally matriculate to the high school.

shall be incorporated into the API: ...
The nationally normed test as augmented ...
The high school exit examination.

...

[S]chools shall be ranked by [API scores] in decile[s] ... [in] three categories: elementary, middle, and high school. The schools shall also be ranked by the value of the API when compared to schools with similar characteristics.³⁹

Characteristics used to determine the similar schools rankings were to include if available, mobility, ethnicity, socioeconomic status, percent of fully credentialed teachers, percent of teachers with emergency credentials, percent of English Language Learners, average class size per grade level, and use of multi-track, year-round programs.⁴⁰ The *Accountability Act* further provided:

Following the annual publication of the API and school rankings by the [Superintendent], the governing board of each school district shall discuss the results of the annual ranking at the next regularly scheduled meeting.⁴¹

The PSAA Advisory Committee

The *Accountability Act* also provided for a PSAA Advisory Committee (*Advisory Committee*) to counsel the Superintendent of Public Instruction [Superintendent] and the *Board* on matters relating to the API, the intervention program, and school awards. Members of the *Advisory Committee* appointed by the *Superintendent* are

³⁹ CEC § 52052(a),(b); CEC § 52056 (a), emphasis added.

⁴⁰ CEC § 52056(a). The original legislation called for the reporting of similar schools API growth. Similar schools rankings were added later.

⁴¹ CEC § 52056(c).

required by statute to be representative and diverse and to serve terms of at most two years without compensation.⁴² The *Advisory Committee* meets several times per year.

In the 1999-2001 and 2001-2003 appointment terms, the *Advisory Committee* has consisted of approximately 8% teachers, 31% administrators, 27% business representatives, 15% college and university professors, 6% government employees, 5% union representatives, and 8% parents representatives. As needed, the *Advisory Committee* has established subcommittees to consider specific issues such as awards criteria, external evaluators, and API calculation rules.

In consultation with the *Department*, the *Advisory Committee* also established the Technical Design Group (*Tech Group*). The *Tech Group* consists of 8 statisticians and measurement specialists from 4 California universities, 2 large school district research and evaluation units, and 2 nonprofit education entities.⁴³ This group was given primary responsibility for developing proposed calculation rules and technical procedures for the API.

API policy was generally adopted by the *Board* via a two-step process. The usual procedure for initial consideration of a new policy was for the Director of the Policy and Evaluation Division and/or a representative of the *Advisory Committee* or *Tech Group* to present an informational

item to the *Board* followed by discussion. The informational item often included a paper with recommendations from the *Advisory Committee* prepared with the assistance of the *Tech Group*. Typically the *Board* made a final policy decision at a subsequent meeting.

An organizational chart of the entities involved in decision-making for the accountability system is presented in Chart 2. A chronological list of *Board* actions involving the API is presented in Chart 3. Chart 4 lists the major policy papers prepared by the *Advisory Committee* and *Tech Group* that informed those policy actions by the *Board*.

Calculating an API

The 1999 API for an elementary school or middle school (grades 2-8) was based on the national percentile rank (NPR) for each valid student *Stanford Test* score in each content area and grade. The following table lists the weights used to calculate an elementary or middle school API.

ELEMENTARY/MIDDLE SCHOOL API WEIGHTS

Performance Bands	Weighting Factors	Content Area	Content Area Weight
80-99 th NPR	1000	Reading	.30
60-79 th NPR	875	Language	.15
40-59 th NPR	700	Spelling	.15
20-39 th NPR	500	Mathematics	.40
1-19 th NPR	200		

The five performance bands used to tabulate student performance were chosen because five performance levels were planned for the California Standards Tests that were to be added to the API later. The chosen weighting factors deliberately gave more weight for moving students up one level in the lowest performance

⁴² CEC § 52052.5.

⁴³ See www.cde.ca.gov/psaa/api/report/tdg for a list of members.

bands than in the higher performance bands.⁴⁴

For example, a student performing at the 25th NPR earns 300 more points than a student performing at the 15th NPR but a student performing at the 85th NPR only earns 125 more points than a student performing at the 75th NPR. The content area weights were chosen so that *language arts* and *mathematics* would be weighted in the ratio 60/40 with reading receiving the greatest emphasis in the *language arts* cluster.⁴⁵

API Calculation Steps

The steps for calculating an elementary or middle school 1999 API were:

1. For a content area test (e.g., *Stanford Test* reading), determine the percent of students across all grades scoring in each NPR performance band (column I).
2. Multiply the percent of students in each performance band by the weighting factor for that band (column II) to obtain the weighted score for each performance band.
3. Sum the weighted scores across performance bands to obtain a total weighted performance bands score.
4. Multiply the total weighted performance bands score for a content area test by its con-

tent area weight (column IV) to obtain a total weighted content area score.

5. Repeat steps 1 through 4 for each of the other content area tests.
6. Sum the total weighted content area scores and round to the nearest whole number to obtain the 1999 API for the school.

An example of a 1999 API calculation for a hypothetical elementary school is presented in Chart 5. The API for a high school was calculated similarly with five *Stanford Test* content areas (reading, language, mathematics, science, social science) each weighted 20%.

Subgroup APIs

In addition to calculating an overall school API, separate APIs are also calculated for each numerically significant ethnic and *SES_D* subgroup. A subgroup is numerically significant if it represents at least 15% of the school enrollment and at least 30 students in that subgroup were tested, *OR* if 100 or more students in that subgroup were tested, whether or not it constitutes 15% of the school enrollment.⁴⁶

Ethnic subgroups for which separate APIs may be calculated include African-American (not of Hispanic origin), American Indian or Alaskan Native, Asian, Filipino, Hispanic or Latino, Pacific Islander, and White (not

⁴⁴ See SBE, Minutes & Agenda Item #23, Nov. 9, 1999, Attachment 1: Advisory Committee, *The 1999 Base Year Academic Performance Index (API)*.

⁴⁵ *Id.*

⁴⁶ CDE, *2000 API Base Report Information Guide*, Jan. 2001, p. 22.

of Hispanic origin).⁴⁷ A student is SES_D if neither parent has earned a high school diploma *OR* if the student participates in the free/reduced price lunch program.⁴⁸ These data are supplied by the student or school by gridding appropriate spaces on the test answer document. A student may be included in the API calculation for both an ethnic subgroup and the SES_D subgroup.

Ranking API Performance

In addition to reporting the overall school API and comparable improvement APIs, state decile⁴⁹ rankings and similar school rankings are also provided to schools for three levels (elementary, middle, and high school).⁵⁰ All of these data are available on the *Department* website and must be reported publicly in the school's annual accountability report card.⁵¹

The overall school API decile ranking indicates where the school's overall API falls in the statewide distribution of school APIs. For example, a school with an overall API decile rank of 8 has performed between the 71st and 80th percentiles statewide.

The similar schools decile rank indicates how the school performed relative to other schools with similar demographic characteristics. A mul-

tipple regression statistical model is used to calculate weights for the 14 demographic characteristics specified in the *Accountability Act*. These weights are then applied to the demographic characteristics of each school to produce a similar schools index. A school's similar schools decile ranking is determined by the placement of its API score among the scores of the 100 California schools with the closest similar schools indices.

Comparison of the overall and similar schools decile ranks can indicate relatively high performing schools that are not achieving as well as they could (e.g., overall rank = 8, similar schools rank = 2). Similarly, relatively low performing schools that are performing much better than other schools with similar challenges (e.g., overall rank = 3, similar schools rank = 9) can be identified.

Calculating Growth Targets

In the California accountability system, schools with at least 100 tested students are expected to achieve two types of API growth each year: (1) overall schoolwide growth, and (2) comparable improvement in the APIs for numerically significant ethnic and SES_D subgroups. The calculations for the growth targets for each of these school growth expectations are described below.

Overall School Growth Targets

An interim API performance standard of 800 was set by the *Board* in 1999. The overall growth target for each

⁴⁷ See API School Reports at www.api.cde.ca.gov.

⁴⁸ *Id.*

⁴⁹ Deciles divide the statewide distribution of API rankings into ten groups each with 10% of the API scores. Deciles range from 1=lowest (bottom 10%) to 10=highest (top 10%).

⁵⁰ CEC § 52056(a).

⁵¹ CEC § 52056(b); CEC § 33126. See section on *Disclosure of Information*, *infra* p. 32, for further discussion on school accountability report cards.

school with an API less than 800 is calculated as follows:

<p>OVERALL GROWTH TARGET =</p> $(800 - \text{API}) \times .05$ <p>(rounded to nearest whole number)</p> <p>OR 1 point,</p> <p>whichever is larger.⁵²</p>

Five percent is the minimum acceptable growth specified in the *Accountability Act*. The calculation was deliberately designed to require schools with APIs further from the interim performance standard of 800 to attain larger overall growth targets.⁵³

For example, using the hypothetical elementary school in Chart 5, the overall growth target is:

$800 - 535 = 265 \times .05 = 13.25$ $\text{rounded} = 13.$

Comparable Improvement Growth Targets

The growth target for demonstrating comparable improvement for all numerically significant ethnic and SES_D subgroups within a school is 80% of the overall growth target rounded to the nearest whole number.⁵⁴ For the

⁵² For schools with at least 100 tested students. Initially, the awards eligibility requirement was also 1 point; the Budget Act of 2001 changed the awards eligibility requirement to a minimum of 5 points. SBE Minutes & Agenda Item #11, Oct. 10, 2001.

⁵³ See SBE, Minutes & Agenda Item #23, Nov. 9, 1999, Attachment 1: Advisory Committee, *The 1999 Base Year Academic Performance Index (API)*.

⁵⁴ If the distance from the subgroup API to 800 is less than the calculated growth target, the distance is used. The subgroup comparable improvement growth target is 4 if the overall growth target is 5 (schools with APIs of 771 to 799) or if the overall API is 800 or more but the

hypothetical elementary school in Chart 5, the subgroup API growth target is:

$13 \times .8 = 10.4$ $\text{rounded} = 10.$
--

Calculating Annual Growth

Annual growth for a school is calculated by subtracting the API for the previous year from the API calculated for the current year. The current year API is referred to as the *growth year API* and the previous year API is referred to as the *base year API*.

<p>ANNUAL SCHOOL GROWTH =</p> $\text{Growth API} - \text{Base API}$ <p>(Current Year API - Previous Year API)</p>

For example, if the hypothetical elementary school in Chart 5 with a 1999 base year API of 535 had a 2000 growth year API of 550, the annual growth would be 15. The school would have met its overall growth target of 13 and would be eligible for awards if the calculated annual growth for each numerically significant subgroup was at least 10.

To calculate the annual school growth, the base and growth year APIs must consist of the same content area tests weighted in the same proportions. When the tests included in the API remain the same,

subgroup API is less than 800. If both the overall API and subgroup APIs are 800 or more, the subgroup must maintain 800 or more. However, to be eligible for awards, a minimum growth of 5 points is required. CDE, *2000 API Base Report*, www.cde.ca.gov, p. 24; SBE Minutes & Agenda Item #11, Oct. 10, 2001.

the current year *growth API* becomes the next year's *base API*.

However, when a new test is added to the API, to provide an appropriate base year API for the following year, the current year API must be recalculated to include the new test with adjusted content area weights. Then there are two APIs for the same year. The growth API is used to measure annual school growth for the current year (relative to the base API for the previous year) and the recalculated API becomes the base API for the following year. The four API reporting cycles from 1999/2000 through 2002/2003 and their included tests are presented graphically in Chart 6.⁵⁵

When the *language arts Standards Test* was added to the base API in 2001 and the *Mathematics Standards Test* and the *High School Exit Exam* were added to the base API in 2002, the revised content area weights were as shown below for elementary/middle schools and high schools.⁵⁶

⁵⁵ When a new base year API is calculated, a scaling adjustment equates the state means for the current year growth API and the new base year API. To date, these scaling adjustments have been very small.

⁵⁶ SBE Minutes & Agenda Item 17, Sept. 5, 2001; Attachment 1: Technical Design Group, *The 2001 Base API: Integrating the California Standards Test for ELA into the API*, July 18, 2001; Attachment 2: Letter from Advisory Committee re Tech Group Recommendations; SBE Minutes & Agenda Item 27, March 6, 2002; Attachment 1: Advisory Committee, *The 2002 Base API: Integrating the California Standards Test in Mathematics into the API*, March, 2002; SBE Minutes & Agenda Item 5, June 26, 2002; Attachment 1: Technical Design Group, *The 2001 Base API: Integrating the California High School Exit Examination (CAHSEE) Results into the API*, April 19, 2002 & *Advisory Committee Recommendations*, May 2002. For the standards tests, the proportion of students in each of the five ordered performance levels (far below basic, below basic, basic, proficient, advanced) is calculated and the same weighting factors used with the 5 *Stanford Test* performance bands are applied. California added the far

REVISED CONTENT AREA WEIGHTS

Content Area	2001 Weights	2002 Weights	Content Area	2001 Weights	2002 Weights
SAT9 Read	.12	.06	SAT9 Read	.08	.03
SAT9 Lang	.06	.03	SAT9 Lang	.08	.03
SAT9 Spell	.06	.03			
Read/ELA Standards	.36	.48	ELA Standards	.24	.35
			HS Exit Exam - ELA	-	.10
LANGUAGE ARTS TOTAL	.60	.60		.40	.51
SAT9 Math	.40	.08	SAT9 Math	.20	.03
Math Standards	-	.32	Math Standards*	-	.18
			HS Exit Exam - Math	-	.05
MATH TOTAL	.40	.40		.20	.26
			SAT9 SSci [†]	.20	-
			History Standards	-	.20
			SAT9 Sci	.20	.03
GRAND TOTAL					
Stanford Test	.64	.70		.76	.82
Standards Tests	.36	.30		.24	.18

[†] In 2003, the 10th & 11th grade history/social science standards tests became the sole social science component.⁵⁷

* General Math, Algebra I, Geometry, Algebra II, Integrated Math 1/2/3, High School Math Standards (Summative) Test.⁵⁸

below basic category to the 4 NAEP performance categories to increase the sensitivity of the Standards Tests to gains by low-achievers. SBE Minutes & Agenda Item #17, Sept. 5, 2001, Attachment 1: Technical Design Group, *The 2001 Base API: Integrating the California Standards Test for English-Language Arts into the API*, July 18, 2001 at 1.

⁵⁷ SBE Minutes & Agenda Item #32, May 30, 2002, Attachment 1: Advisory Committee, *The 2002 Base Academic Performance Index (API): Changes in the High School Social Science Indicator*, May 2002.

⁵⁸ Math Standards Tests in grades 8-11 are administered according to course enrollment as follows:

Math Standards Test Administered	Course Taken This Yr	One Math Standards Test per Grade Based on Course Taken			
		Grade 8	Grade 9	Grade 10	Grade 11
General Math	Pre-Alg, 1 st yr of 2-yr Alg I, gen math, or no math	✓	✓		
Algebra I	Algebra I or 2 nd yr of 2-yr Alg I	✓*	✓	✓	✓
Geometry	Geometry	✓	✓	✓	✓
Algebra II	Algebra II	✓	✓	✓	✓
Integrated 1-3	Integrated 1-3	✓ ¹	✓ ²	✓ ³	✓
HS Math Stand	Completed Alg II or Int 3		✓	✓	✓

* Cells in blue = recommended sequence for full mastery of state math standards & completion of HS Math Standards Test by grade 11.

Adapted from SBE Minutes & Agenda Item 27, March 6, 2002; Attachment 1: Advisory Committee, *The 2002 Base API: Integrating the California Standards Test in Mathematics into the API*, March, 2002 at 7.

When the *language arts* and mathematics Standards Tests and the *High School Exit Exam* were added to the school API calculations, the 200-1000 scale range, the 800 interim performance target, and the weights for each performance band as applied to the corresponding performance categories remained the same.⁵⁹ The 60/40 ratio of *language arts* to math in elementary and middle schools was retained but the relative weights for the content areas at the high school level were changed in 2002. In addition, for the 2002 base API, the weight of the norm-referenced test was decreased to keep the API more stable when the *Stanford Test* is replaced in 2003 by the shorter, survey version of the *California Test*.⁶⁰

With the addition of the Standards Tests, the major changes in the calculation of the API included revision of the weights within content areas and addition of a scaling factor.⁶¹ Within each content area at the elementary and middle school levels, 80% of the weight was assigned to

⁵⁹ High school students in grades 10 & 11 who took no Math Standards Test were assigned a 200 point credit. Since this represented the lowest possible point value, it provided an incentive for schools to encourage enrollment in recommended math courses. SBE Minutes & Agenda Item 27, March 6, 2002; Attachment 1: Advisory Committee, *The 2002 Base API: Integrating the California Standards Test in Mathematics into the API*, March, 2002 at 13.

⁶⁰ SBE Minutes & Agenda Item #5, Jan. 8, 2003, Attachment 1: Technical Design Group, *Revisions in the 2002 Base Academic Performance Index (API)*.

⁶¹ Because the *High School Exit Exam* results were reported as pass or fail, the performance weights were 1000 points for students who passed in grades 10-12 and 200 points for students who failed in grade 10. The API component total was the average of all student points across grades. SBE Minutes & Agenda Item #5, June 26, 2002, Attachment 1: *The 2002 Base Academic Performance Index (API): Integrating the California High School Exit Examination (CAHSEE) Results into the API*, April 19, 2002 at 8-9. See also footnote 55.

the Standards Test and 20% to the *Stanford Test*. At the high school level in the content areas of *language arts* and mathematics, 69% of the weight was assigned to the *Standards Test*, 19% to the *High School Exit Exam* and 12% to the *Stanford Test*.⁶²

A scaling factor was applied to the resulting API scale for each level (elementary, middle, high school) to equalize the state mean on the new scale to its previous value (i.e., CA 2002 Growth API high school mean [including *Stanford Tests* and *Language Arts Standards Tests*] = CA 2002 Base API high school mean [including *Stanford Tests*, *Language Arts* and *Mathematics Standards Tests*, and the *High School Exit Exam*]).⁶³

Chart 7 graphically demonstrates the purpose for using a scaling factor. A sample 2002 base API calculation for a hypothetical high school is presented in Chart 8.

With the addition of the *Language Arts Standards Test* in 2001, the *Stanford Test* weight decreased from 100% to 64% at the elementary and middle school levels and from 100% to 76% at the high school level. When the *Mathematics Standards Tests*, the *High School Exit Exam* and the *High School History Stan-*

⁶² SBE Minutes & Agenda Item #5, Jan. 8, 2003, Attachment 1: Technical Design Group, *Revisions in the 2002 Base Academic Performance Index (API)*.

⁶³ SBE Minutes & Agenda Item 17, Sept. 5, 2001; Attachment 1: Technical Design Group, *The 2001 Base API: Integrating the California Standards Test for ELA into the API*, July 18, 2001; SBE Minutes & Agenda Item 27, March 6, 2002; Attachment 1: Advisory Committee, *The 2002 Base API: Integrating the California Standards Test in Mathematics into the API*, March, 2002.

dards Tests were added to the API in 2002, the weight for the *Stanford Test* decreased from 64% to 20% at the elementary and middle school levels and from 76% to 12% at the high school level while the weights for the Standards Tests increased to 80% and 88%, respectively.

Additional reductions in the weight of the norm-referenced test will occur when the other Social Science and Science Standards Tests are included in the API in future years. The *Department's* proposed six-year plan for the introduction of additional indicators into the API is summarized in Chart 9.

Summary of the API

The API, mandated by the *Accountability Act*, is a summary school level measure of student academic performance in the content areas of *language arts*, mathematics, science and history/social science.

The API

- Provides scores on a scale ranging from 200 to 1000;
- Ranks schools on a scale ranging from 1 to 10;
- Sets a statewide performance target of 800;
- Assigns schools specific growth targets for future improvement; and
- Provides comparisons between schools with similar characteristics.⁶⁴

Each year, the *Board* reviews and approves the proposed components, weights, and calculation rules for the API.⁶⁵

API Awards

In the summer of 2000, schools meeting their API overall and comparable improvement growth targets were eligible for two types of awards: The Governor's Performance Award Program (GPA or *Awards Program*) and the Certificated Staff Performance Incentive Award (Assembly Bill 1114). The former awards were based on a fixed dollar amount per student while the latter were awards to teachers. Schools and their staff were also eligible for a one-time school site employee performance bonus (Senate Bill 1667). A comparison of the characteristics and criteria for these awards as recommended by the *Advisory Committee* and approved by the *Board* in July 2000 is presented in Chart 10.⁶⁶

The *Board* adopted Emergency Regulations for API-based awards programs with minor revisions in October 2000.⁶⁷ API permanent regulations were approved in March 2001.⁶⁸

⁶⁴ Adapted from CDE, Office of Policy and Evaluation, overhead #4, June 7-8, 2000 SBE Final Minutes, Attachment 5, p. 3.

⁶⁵ SBE, Minutes & Agenda Item #11, July 12, 2000; SBE, Minutes & Agenda Item #7, Sept. 5, 2001; SBE, Minutes & Agenda Item #27, March 6, 2002; SBE, Minutes & Agenda Item #5, June 6, 2002; SBE, Minutes & Agenda Item #5, Jan. 8, 2003. See also Chart 3.

⁶⁶ SBE, Minutes & Agenda Item #11, July 12, 2000.

⁶⁷ SBE, Minutes & Agenda Item #12, Oct. 11, 2000. Subsequently, a public hearing was held and the Emergency Regulations (CCR § 1031 et seq.) were re-adopted with minor revisions. SBE, Minutes & Agenda Items #35 & #36, Jan. 11, 2001.

⁶⁸ SBE Minutes & Agenda Item #21, March 7, 2001.

The *Accountability Act* directed the *Board* to establish the *Awards Program* to:

provide monetary and nonmonetary awards to schools that meet or exceed [their API overall growth targets] and demonstrate comparable improvement in academic achievement by all numerically significant ethnic and SES_D groups within schools.⁶⁹

To be eligible for an award, a school must satisfy the following:⁷⁰

Awards Criteria

- overall API growth target
- comparable improvement growth target for each numerically significant subgroup
- participation rate 90-95%
- 85% tested in each content area
- parent exemptions < 10%
- no reports of adult testing irregularities.

To be awards eligible, the *Board* adopted minimum participation rates of 95% for elementary and middle schools and 90% for high schools.⁷¹ This action was consistent with the

⁶⁹ CEC § 52057(a).

⁷⁰ California Code of Regulations (CCR) § 1031 et seq.; SBE Minutes & Agenda Items #16 & #17, Feb. 7, 2001; Advisory Committee, *Summary of Policy Issues Related to the 2001 Growth API & Issue Papers*, CDE, Feb. 2001, Attachments 1-6. The minimum growth for awards eligibility is 5 points. *Supra* note 54.

⁷¹ CCR § 1032(h). The *Board* intends to increase the participation rate for high schools to 95% in the future. SBE, Minutes & Agenda Item #11, July 12-13, 2000. The decision to adopt a lower participation rate requirement for high schools was based on significantly lower 1999 participation rates for high schools than for middle or elementary schools (80% of elementary schools, 79% of middle schools, and 27% of high schools receiving an API in 1999 met the 95% criterion). *Id.* at Attachment 8.

previously adopted *API Framework* that stated:

The API should reflect the achievement and the growth of all students ... All schools should apply common standards of pupil participation in assessments (i.e., inclusion criteria) to enable valid comparisons among schools.⁷²

A school's participation rate is calculated as:

$$\text{SCHOOL PARTICIPATION RATE} = \frac{\text{number of tested students}}{\text{enrollment} - \text{special education exemptions.}}^{73}$$

With advice from the Accountability Advisory Committee, the *Board* established a sliding scale of per pupil award amounts such that schools with API gains exceeding their growth targets by the largest amounts received the largest per student awards. Schools exceeding their API growth targets were rank ordered by size of gain and the award money was distributed according to specified percentages of the total funds allocated until all available funds were exhausted. All schools that met or exceeded their growth targets and demonstrated comparable improvement were given certificates and a special emblem to use on their letterhead during the following year.⁷⁴

On behalf of a school eligible for the *Awards Program*, its district govern-

⁷² API Framework, *supra* note 32 at 3.

⁷³ *Id.* Parental waivers were also subtracted from the denominator for the initial two API growth cycles. The *Board* changed this policy in January 2001. SBE, Minutes & Agenda Item #36, Jan. 11, 2001.

⁷⁴ See www.cde.ca.gov/ope/awards/govperfl/.

ing board may request a three-year waiver of state statutory or regulatory provisions providing that required instructional time does not decrease nor state costs increase. Such waivers may include maximum flexibility in the allocation of expenditures for the successful school.⁷⁵

Monetary awards were given in 2000 and 2001 and then discontinued due to budget deficits. Authority for non-monetary awards has continued.

Interventions

The 1999 *Accountability Act* established the Immediate Intervention/Underperforming Schools Program (II/USP or *Intervention Program*).⁷⁶ Schools scoring in the lowest five deciles on the API calculated from their *Stanford Test* scores were eligible to apply the first year. In subsequent years, the additional criteria of failing to meet overall and comparable improvement growth targets was added. The 430 schools specified by the statute to be selected each year were to be spread across deciles, school levels and geographic locations.⁷⁷

Cohort 1 schools were selected in the fall of 1999, Cohort 2 schools in the fall of 2000, and Cohort 3 schools in the fall of 2001. The number of eligible schools, the number and percent of eligible schools that applied, and the number and percent of applicant schools selected

⁷⁵ CEC § 52057(d),(e).

⁷⁶ CEC § 52053.

⁷⁷ *Id.* at (a)-(h).

for the *Intervention Program* are shown below.⁷⁸

COHORT	ELIGIBLE	APPLIED	SELECTED
	3145	1423(45%)	430(14%)
	936	528(56%)	430(46%)
	1266	751(59%)	430(34%)

Participating schools were awarded state planning or federal implementation grants of at least \$50,000. The school board was required to "appoint a broad-based schoolsite and community team, consisting of a majority of nonschoolsite personnel" and to select an external evaluator from a *Board* approved list.⁷⁹ The school teams and external evaluators were required to collaborate to identify weaknesses contributing to the school's poor test performance and develop an action plan to implement recommendations for improvement.⁸⁰

To the extent data was available for the school, two-year, short-term objectives for improvement were to include consideration of *Stanford Test* scores, graduation rates, and attendance rates for students and school personnel. When developing their action plans, schools were also required to review crime statistics, consider school conditions identified in their school accountability report cards, and consult with collective

⁷⁸ Farr, B. & O'Day, J., *Evaluation Study of the Immediate Intervention/Underperforming Schools Program and the High Achieving/Improving Schools Program of the Public Schools Accountability Act of 1999, Phase I Report*, American Institutes for Research, June 30, 2002, p. 10.

⁷⁹ In June 1999, the *Board* approved the minimum qualifications, application process, and standards and criteria for the external evaluators. SBE, Minutes & Agenda Item #18, June 10, 1999.

⁸⁰ CEC § 52053(f),(l); § 5053.5; § 52054(a),(c).

bargaining representatives where applicable.⁸¹

At the end of the initial planning year, the school submitted its action plan and request for funding to the *Superintendent*. The *Superintendent* provided a recommendation to the *Board* which made a funding decision and determined whether any requests for rule waivers would be granted. Funded schools received \$200 per student (or a minimum of \$50,000) for two years and were required to match the amount of the state grant with new or existing revenues.⁸² A participating school that meets or exceeds its growth targets each year is eligible for awards and exits the program at the end of the two years of funding.⁸³

Participating schools that do not meet their API growth targets during the first year of funding are required to hold a public hearing to communicate the lack of progress to the community and to choose from a range of interventions to further the action plan. If the school fails to meet its growth targets after a second year of funding but has demonstrated significant growth, a third year of funding may be provided, followed by state-monitoring and program exit after 2 consecutive years of significant API growth.⁸⁴

A participating school that has not met its growth targets or demonstrated significant growth after 2 years is deemed a state-monitored

school. The *Superintendent*, with *Board* approval, may take over the management of the school or may assign an intervention team.⁸⁵ Any action taken must be accompanied by:

specific findings by the [*Superintendent & Board*] that the action is directly related to the *identified causes for continued failure by a school to meet its performance goals*.⁸⁶

Summary results for Cohort 1 are presented in Table 4. Of the 430 participating schools, 68% met their targets or made at least some growth, 16% received a third year of funding, and only 5% failed to make any growth and are subject to State sanctions.

Evaluation Reports

For all *Intervention Program* participating schools in their jurisdiction, school districts are required to submit an annual evaluation to the *Superintendent*. The district evaluation must include an analysis of the impact, costs and benefits of the program, the status of participating schools relative to their growth targets, and reasons schools have or have not met their growth targets.⁸⁷

In addition, the *Accountability Act* specified the selection of an external auditor to evaluate the intervention and award programs.⁸⁸ The American Institutes for Research (AIR) was selected by the *Board* as the

⁸¹ CEC § 52054(d),(h).

⁸² CEC § 52054(i),(j); § 52054.5.

⁸³ CEC § 52055.5(a).

⁸⁴ CEC § 52055; § 52055.5 et seq.

⁸⁵ CEC § 52055.5 et seq. The statute also affords certain protections to principals targeted for replacement.

⁸⁶ CEC § 52055.5(j), emphasis added.

⁸⁷ CEC § 52058(a).

⁸⁸ CEC § 52058(b).

external evaluator for the two-year study. The external evaluator produced an interim Phase I report including API and *Stanford Test* results through 2001.⁸⁹

The external evaluator concluded:

□ [O]n balance, both Cohort 1 and Cohort 2 [*Intervention Program*] schools appear to have experienced faster achievement growth than the comparison schools, subsequent to their participation in the program.

□ [T]here is some evidence that the difference in achievement growth between [*Intervention Program*] and comparison schools is most pronounced during the initial [planning year of funding].⁹⁰

Sample graphs of Cohort 1 and comparison elementary schools' API, *Stanford Mathematics* and *Stanford Reading* performance from 1998 through 2001 for state-funded schools are presented in Chart 12c. These graphs summarize data for two years prior to *Intervention Program* funding (1998 & 1999) and the following two years of program participation (2000 & 2001).

Based on qualitative analyses and impressions from an initial set of site visits, the external evaluator also concluded:

[I]t appears that ... [the *Intervention Program*] is largely fulfilling its intent. That is, it seemed the program unfolded as envisioned:

- It provided focus for schools' planning efforts;
- it resulted in the implementation of improvement strategies;

- it provided much needed resources; and,
- it appeared that there were some positive outcomes – both in terms of school culture and early reports of student achievement.

In particular, the role of [the *Intervention Program*] in fostering a focus on improvement strategies seemed most pervasive. All of the researchers reported that interviewees state that if nothing else, the [*Intervention Program*] initiative focused their efforts and led them to consider what was working in their schools as well as what needed improvement.⁹¹

The final report will extend the data analyses to include 2002, finish site visits for all selected schools, survey participants and stakeholders, and make recommendations for program improvement. The final report will also systematically evaluate positive and negative unintended consequences of the *Intervention Program*.

Alternative Accountability System

The *Accountability Act* provided for the creation of an alternative accountability system for schools with less than 100 valid test scores or serving special populations. In July 2000, the *Board* approved the recommendations of the *Advisory Committee* for a three-part alternative accountability framework to be implemented over three years.⁹² Included were: (1) traditional schools with 11-99 valid test scores who receive an asterisked API to indicate larger statistical uncertainty; (2) special education schools and centers; and (3) alternative schools serving

⁸⁹ AIR Evaluation Report, *supra* note 78.

⁹⁰ *Id.* at v.

⁹¹ *Id.* at vi.

⁹² SBE, Minutes & Agenda Item #11, July 12, 2000.

mostly high-risk students (e.g., county court schools for juvenile offenders) and very small schools with less than 11 valid test scores.

EVOLUTION OF THE API

The API is a work in progress. Since its inception in 1999, the API has evolved with experience and the availability of more targeted Standards Tests. The *API Framework* adopted by the *Board* specifically provided that:

the API should transition to upgrades in statewide assessments as they become aligned to statewide content and performance standards. Achievement of state education goals—state content and performance standards—must be the framework for the focus on student performance.

...

The API should be flexible to accommodate incorporation of future indicators or components and should evolve in an orderly fashion as additional indicators become available and are incorporated over time. Within each overlapping two-year API cycle, all component indicators should be the same, with common definitions from year to year.⁹³

From the beginning, the intent was to hold schools accountable for teaching the state content standards. But it has taken time to develop valid and reliable standards tests (as required by statute) that are tailored specifically to the California content standards and linked to California performance standards.⁹⁴ In the interim, the *Board* chose to base the API calculation on a valid and reliable na-

tionally standardized achievement test (the *Stanford Test*) with substantial overlap with the California content standards.⁹⁵ However, the methodology approved by the *Board* for calculating the initial API based only on the *Stanford Test* results allowed for the later inclusion of California Standards Tests.

Addition of Standards Tests

The off-the-shelf achievement test was augmented with additional items tied specifically to state standards not already measured in that test. When a large enough item pool had been developed and tested, separate California Standards Tests were developed and are gradually being introduced into the API. Initially, the Standards Tests were not included in the API because they were not yet fully developed, their validity and reliability as required by law had not yet been fully demonstrated, and performance standards had not yet been established.⁹⁶

When a Standards Test for a content area is added to the API, the corresponding nationally-normed achievement test in that content area receives substantially less weight. When all the California Standards Tests have been added to the API, they will collectively constitute most of the weight of the academic tests in the API. However, some weight will remain on the content area subtests of the nationally-normed achievement test to provide more accurate measurement of low-performing stu-

⁹³ API Framework, *supra* note 32 at 2-3.

⁹⁴ The 1999 Base Year API, *supra* note 35 at 4.

⁹⁵ SBE, Minutes & Agenda Item #23, Nov. 9, 1999.

⁹⁶ API Framework, *supra* note 32 at 6.

dents and to retain the ability to make national comparisons.

Once all new components have been added to the API, separate base and growth APIs will no longer be necessary. There will only be a single, annual API data release and meaningful year-to-year comparisons of API scores will be possible.⁹⁷

Future Plans

Chart 9 summarizes the indicators included in the API so far and the *Department's* plan for inclusion of additional indicators in the future that was reviewed by the *Board* at its April 24, 2002 meeting. As indicated in Chart 9, the *Language Arts Standards Test* was added to the base API in 2001 and the *Mathematics Standards Test*, the high school *History Standards Test*, and the *High School Exit Exam* were included in the 2002 base API. Integration of the *Mathematics Standards Tests* and the *High School Exit Exam* into the API presented unique technical challenges because they are not universal (all students at all grade levels do not take these tests).⁹⁸

The remaining *Social Science and Science Standards Tests* will be added incrementally by grade beginning with the 2003 base API. In 2003, the *Stanford Test* will be replaced by the *California Test*.⁹⁹ Stability of the API scale will be maintained by equating the *California*

⁹⁷ *Id.* at 3.

⁹⁸ *Id.* at 2.

⁹⁹ SBE Minutes & Agenda Items #1 & #9, April 24, 2002, Attachment 1: CDE, *The Academic Performance Index (API): A Six-Year Plan for Development (2001-2006)*, April 25, 2002, at 2-3.

Test to the *Stanford Test* and by the reduction in weight of the norm-referenced component to 20% at the elementary and middle school levels and 12% at the high school level.¹⁰⁰

In addition to the California standards tests being added to the API, the state also plans to consider adding other measures including the attendance and graduation rates specified in the *Accountability Act*.¹⁰¹ Caution in the inclusion of nontest indicators was contemplated by the *Accountability Act*. It states:

Before including high school graduation rates and attendance rates in the [API], the [Superintendent] shall determine the extent to which the data are currently reported to the state and the accuracy of the data.¹⁰²

The addition of graduation and attendance rates is problematic because the state does not currently collect this information for all schools and the information it does collect may not be based on consistent definitions or accurate calculations of these statistics. It is difficult for schools to determine which students have actually dropped out because many students leave schools without requesting transfer of their records or providing any information about their future plans. Students are not required to do so and some families may view this information as private. Further, special data collection pro-

¹⁰⁰ SBE Minutes & Agenda Item #5, Jan. 8, 2003, Attachment 1: 2003, Attachment 1: Technical Design Group, *Revisions in the 2002 Base Academic Performance Index (API)*.

¹⁰¹ *Id.* at 7-9.

¹⁰² CEC § 52052(a).

cedures would be needed to add staff attendance data.¹⁰³

Currently, participation in the California School Information Services (CSIS) database, that collects graduation and student attendance data, is voluntary and non-universal. The *Department* has estimated the start-up and continuing costs for the expansion of this database to include all schools and all measures to be \$60 million over four years.¹⁰⁴

Moreover, including graduation rates and attendance rates in the API presents the following additional technical issues:

- Establishing performance standards for these rates would be more arbitrary and open to debate than for test results.
- Between-school variability may be small for attendance rates resulting in their having little effect on the API and creating little room or incentive for improvement.
- Inclusion of these rates is inconsistent with the API Guiding Principles which call for student outcome rather than process indicators. In particular, staff attendance is not student-based.¹⁰⁵

Similar problems can be expected with any attempts to include other input (process) variables (e.g., percent of credentialed teachers or ratio of textbooks to students) as indicators in the API. Nonetheless, based on the *Board-adopted API Framework* stating that:

¹⁰³ *Id.* at 3.

¹⁰⁴ CDE, *Establishing School-Level Graduation and Attendance Rates for Implementation in School Accountability, Addendum on Estimated Costs*, report to the Governor and Legislature, Nov. 15, 1999.

¹⁰⁵ *API Six-Year Development Plan*, *supra* note 99 at 4.

[a]s California transitions to the implementation of the [*Information System*], the API should accommodate indicators emanating from [it],¹⁰⁶

the *Department* and the *Board* plan to continue studying the feasibility and appropriateness of including nontest measures in the API.

The state is also studying the following proposals for modifying the API:

- Adding the California English Language Development Test (CELDT) used to measure the progress of English language learners in learning English.
- Adding the California Alternate Performance Assessment (CAPA) used to assess severely disabled special education students whose IEPs exempt them from academic standardized tests.
- Improving the reliability of estimates of school growth by pooling data across years in multi-year accountability cycles. This would be particularly helpful for small schools (11-99 tested students) that currently receive an asterisked API but no annual growth targets and are not eligible for state awards.
- Introducing a *value added* measure of individual student longitudinal growth. California presently does not have the state-level database required for such calculations.
- Considering whether high-scoring student subgroups should be required to meet the same comparable improvement growth target as low-scoring subgroups. To narrow the gap in achievement between subgroups, low-scoring subgroups need to make greater gains.
- Collecting high school math courses previously taken on the test answer document to identify students who have completed the High School Mathematics

¹⁰⁶ *API Framework*, *supra* note 32 at 3.

Summative Standards Test (normally administered in grade 11) early so they can be eliminated from the calculation of non-tested students.¹⁰⁷

RESPONSE TO THE RUSSELL REPORT

The report of plaintiffs' expert, Michael Russell, examines the California testing and accountability system described above. He critiques the system with respect to choice of indicators, validity of test data, policy decisions, error in the system, consequential validity, lessons from other states, public opinion, and teacher variables.

Unfortunately, the Russell Report is heavy on conjecture, short on supporting evidence and dismissive of contradictory information. Specific examples are discussed in the following sections.

Choice of Indicators

Russell assumes that the purpose of the California accountability system should be to:

accurately and sufficiently notify the State of whether students receive essentials required for learning,

and based on that assumption, faults the API for not doing so, claiming that his preferred input-based alternatives would produce superior results.¹⁰⁸

¹⁰⁷ *Id.* at 4-6; SBE Minutes & Agenda Item 27, March 6, 2002; Attachment 1: Advisory Committee, *The 2002 Base API: Integrating the California Standards Test in Mathematics into the API*, March, 2002 at 13.

¹⁰⁸ Russell, M. *Expert Report Submitted for: Eliezer Williams vs. State of California*, October 2002, p. iv-xxi.

Russell contends that the current accountability system should be replaced by a new system focusing on input variables such as student/teacher ratios, percent of teachers with emergency credentials, ratio of library books to students, overcrowding of facilities, type of school calendar, professional development opportunities, etc. Student test scores would be only one of at least 9 suggested outcomes that *might* also be included in his system.¹⁰⁹

However, while making the argument that the API should measure inputs rather than student outcomes, Russell appears to dismiss the stated intent of the *Accountability Act*, provides *no* evidence that his system would promote greater student learning, ignores evidence that the current system has had beneficial effects on the academic achievement of poor and minority students, appears to have dismissed the role of inputs in the current accountability system, and ignores contrary evidence from an unsuccessful, court-imposed inputs model implemented in Kansas City, Missouri.

Stated Intent of the Accountability Act

Russell claims that the purpose of California's accountability system is unclear. But this is not true. The purpose is set forth clearly and succinctly in California law.

¹⁰⁹ *Id.* at xvii-xix.

The *Accountability Act* states that:

the purpose of the California public school system is to provide for the *academic development* of each pupil

and that the purpose of the accountability system is

to hold each of the state's public schools accountable for the *academic progress and achievement* of its pupils within the resources available to schools.¹¹⁰

Implementing the expressed intent of this legislation, the current California academic accountability system focuses primarily on student outcomes – that is, it seeks evidence that students are learning the knowledge and skills set forth in the Board-adopted California content standards for *language arts* (reading and writing), mathematics, science and history/social science – and specifies that such evidence be weighted heavily in evaluating school performance (at least 60% of the API).¹¹¹

When schools are successful in this endeavor, they are eligible for awards and are allowed greater freedom (waivers) from state rules.¹¹² When schools are unsuccessful, they are eligible to apply for intervention grants (430 schools selected per year),¹¹³ and if the school continues to be unsuccessful for 3 years in the *Intervention Program*, as

¹¹⁰ CEC § 52050.5, emphasis added.

¹¹¹ PSAA, CEC § 52052 (1999).

¹¹² CEC § 52056 & § 52057.

¹¹³ If enough schools do not volunteer for the *Intervention Program*, the *Department* can designate schools for participation. Of the approximately 7200 schools with APIs, about 1200 were eligible for the *Intervention Program* the first year. With funding available for 430 per year, all eligible schools could participate within 4 years. Padia deposition at 64, 330.

a last resort, the state intervenes directly.¹¹⁴

The *API Framework*, drafted by the *Advisory Committee* and adopted by the *Board*, states:

As important as it is to focus on the many central features of schooling that might be considered as indicators (e.g., teachers, instructional resources, curriculum, an school organization), the primary emphasis of the API is student performance.¹¹⁵

Improved Ethnic & SES_D Subgroup Achievement

Russell asserts that:

[T]he [*Accountability Act's*] single-minded focus on student outcomes as measured by standardized tests fails to adequately prevent, detect or deter gross disparities in education.¹¹⁶

But focusing on outcomes is mandated by statute. Educators and legislators in California, like those in many states, have determined that focusing on outcomes is preferable to continuing to pursue the failed input-based policies of the past.¹¹⁷

Moreover, the available evidence suggests that the performance of historically disadvantaged subgroups (African-Americans, Hispanics, *SES_D* students, English Language Learners) has improved since the inception of the *Assessment Act* and API programs and that the gap in per-

¹¹⁴ CEC § 52053 - § 52055.5.

¹¹⁵ API Framework, *supra* note 32 at 2.

¹¹⁶ Russell Report, *supra* note 108 at vi.

¹¹⁷ Araki, S. & McCully, Co-chairs, Rewards & Interventions Advisory Committee, *Steering by Results*, CDE, 1997, p. 2.

formance between these subgroups and majority group students has narrowed.

Statewide Data. Table 2a summarizes statewide *Stanford Test* results for reading and mathematics from 1998 through 2002 for all students, ELLs beyond 12 months in the district, socio-economically disadvantaged students, African-Americans (AA), and Hispanics (H) in selected grades. In each cell, reading results are presented first followed by mathematics results in parentheses.

The national percentile rank (NPR) for the average student score is highest in the elementary grades and higher in math than in reading for all groups. NPR gains for the average student score over the five-year period 1998-2002 for ALL and ELLs (or increase in percent of students above the 50th NPR over the four-year period 1999-2002 for SES_D, AA and H) were as follows.

Read (Math)	ALL	ELL	SES _D	AA	H
	13(19)	16(23)	10(14)	12(14)	13(16)
	10(19)	9(17)	9(14)	9(15)	9(18)
	5(9)	5(8)	3(7)	4(7)	5(9)
	1(5)	1(3)	0(0)	2(3)	1(4)

All groups made gains over the past several years, most notably in the elementary grades. African-Americans made similar gains to Hispanics in reading but lower gains in mathematics. ELLs generally gained more than SES_D in both content areas. In 2002 in grade 2, average performance in all groups in mathematics was about at the national average with reading performance 5 to 15 percent lower. High school performance was relatively unchanged

across the board but may improve when cohorts that received standards-based instruction throughout elementary and middle school reach the high school level.

Table 2b presents the corresponding statewide summary data for the *Language Arts* and *Mathematics Standards Tests*. However, data across years cannot be as readily compared because data for the initial years was reported as percent correct and then changed to percent proficient once performance standards had been established. Nonetheless, comparisons of percent correct for the two-year period from 1999 to 2001 for All Students, ELLs and SES_D indicate the following changes.

Lang Arts (Math)	ALL	ELL	SES _D
	4(12)	5(12)	5(12)
	4(13)	3(11)	4(13)
	1(11)	2(7)	2(9)
	1(-7)	-1(-4)	-1(-5)

Again, gains were greatest at the elementary level. High school scores decreased during this period but that may have been due to changes in the difficulty of the test because the reported results do not reflect equating of test forms across years. Nonetheless, in comparing groups on the same metric, ELL and SES_D students made about the same amount of gain on average on the language arts and mathematics Standards Tests as all students statewide. Moreover, a study presented to the Accountability Advisory Committee indicated that schools with a higher proportion of English Language Learners had a slightly higher projected growth on the API

than schools with a lower proportion of English Language Learners.¹¹⁸

Based on currently-available data, Charts 14a and 14b summarize the comparison of gains across years for these groups on the *Stanford Test* and Standards Tests. When additional data on changes in percent proficient or above become available in future years, the relative gains for these groups on that more meaningful metric can be determined. By that time, nearly all of the API weight will be on the Standards Tests.

Data for Individual Districts & Schools. Table 3a and Charts 16a and 16b compare the *Stanford Test* reading and mathematics performance of several school districts with significant numbers of ELL and SES_D students in selected grades from 1998/99 to 2002. Growth in performance in terms of the NPR for the average California student in each group is provided for comparison.

In the elementary grades, two of the districts demonstrated growth for the ELL and SES_D groups that generally exceeded the state average while two had growth below the state average. These patterns continued on a smaller scale in middle school and high school grades with the poorer performing districts posting negative gains in grade 10. Clearly, there is a substantial difference in success among districts serving ELL and SES_D populations. The two poorer performing districts had greater gains in mathematics but still generally lagged behind the more suc-

cessful districts. Again, the poorer performing districts had negative gains in grade 10.

In the more successful districts, ELLs generally made gains equal to or greater than those made by all students in the district. In the poorer performing districts, the reverse was true. By 2002 in the elementary grades, the gap between the performance of ELL and SES_D subgroups and all students had generally narrowed in the more successful districts while widening in the poorer performing districts.

Table 3b presents similar subgroup comparison data for selected schools attended by named plaintiffs.¹¹⁹ Similar to the districts compared above, these schools also demonstrate different patterns of success with ELL and SES_D students. At the elementary grades, subgroup students demonstrated substantial growth that was generally comparable to or greater than that for all students in Cahuenga and Coronado. At Edison-McNair, growth in grade 4 was much smaller.

In middle school, Bunche generally performed better than the other two middle schools, although ELL performance was volatile due to small numbers. At the high school level, Dorsey outperformed Crenshaw and generally exceeded state average growth in reading and mathematics. These data demonstrate that schools, like districts, can have dif-

¹¹⁸ SFUSD et al. v. SBE, Padia deposition, May 2, 2000, p. 79.

¹¹⁹ These schools were among those listed by plaintiffs as having teacher quality problems. See section on *Data for Schools Attended by Named Plaintiffs*, *infra* p. 75.

ferential success with ELL and SES_D subgroups.

Role of Inputs in the Current Accountability System

Disclosure of Information. Russell asserts:

A truly comprehensive accountability system would ask schools to describe the programs and practices they have in place, the appropriateness of these programs and practices given specific context and background indicators, and the effect these programs have on a variety of student outcomes.¹²⁰

But Russell also admitted in deposition that no state has what he considers to be a “truly exemplary accountability system.”¹²¹ In addition, California actually requires schools to do many of the things Russell suggests.

According to state law, all California public schools must produce an annual report card containing, among other things, input data and performance measures disaggregated for ethnic and SES_D subgroups.¹²² The school accountability report card statute provides:

The school accountability report card shall provide data by which parents can make meaningful comparisons between public schools

The school accountability report card shall include, but is not limited to, assessment of the following school conditions [for the most recent three-year period]:

- (1) Pupil [*Stanford Test*, Standards Tests, & SAT] achievement by grade level,
- (2) Progress toward reducing dropout rates, . . . , and the graduation rate,
- (3) Estimated expenditures per pupil and types of services funded.
- (4) Progress toward reducing class sizes and teaching loads,
- (5) The total number of the school’s fully credentialed teachers, [teachers with] emergency credentials, [teachers] without credentials, and any assignment of teachers outside their subject areas of competence
- (6) Quality and currency of textbooks and other instructional materials, including whether textbooks and other materials meet state standards . . . , and the ratio of textbooks per pupil and the year the textbooks were adopted.
- ...
- (8) Availability of qualified substitute teachers.
- (9) Safety, cleanliness, and adequacy of school facilities.
- (10) Adequacy of teacher evaluations and opportunities for professional improvement,

...

(18) The [API], including the disaggregation of subgroups [] and the decile rankings and a comparison of schools.

...

It is the intent of the Legislature that schools make a concerted effort to notify parents of the purpose of the school accountability report cards, as described in this section, and ensure that all parents receive a copy of the report card; to ensure that the report cards are easy to read and understandable by parents; to ensure that local educational agencies with access to the Internet make available current copies of the report card through the internet; and to ensure that administrators and teachers are available to answer any questions regarding the report cards.¹²³

¹²⁰ Russell Report, *supra* note 108 at xvii.

¹²¹ Deposition of M. Russell, Los Angeles, CA, January 16-17, p. 134, line 17 – p. 135, line 3.

¹²² CEC § 35256.

¹²³ CEC § 33126.

State statute also requires the *Department* to develop and the *Board* to adopt a template:

to simplify the process for completing the school accountability report card and make [it] more meaningful to the public. . . . When the template for a school is completed, it should enable [comparisons of] how local schools compare to other schools within that district [and the state].¹²⁴

The *Department* is further instructed to provide schools with standard definitions for school conditions required to be included in the school accountability report card and to maintain links to the internet locations of those schools that post their report cards on the internet. The statute states:

Definitions shall enable schools to furnish contextual or comparative information to assist the public in understanding the information in relation to the performance of other schools.

...
The [*Department*] shall monitor the compliance of local educational agencies with the requirements to prepare and to distribute school accountability report cards."

...
The [*Superintendent*] shall additionally review, and the [*Board*] shall consider, any empirical research data that becomes available concerning barriers to equal opportunities to succeed educationally for all California pupils, regardless of socioeconomic background. Upon obtaining this information, the board shall evaluate whether there is any need to revise the school accountability report card.

...
The State Allocation Board, in cooperation with the [*Superintendent*], shall develop and maintain an automated school facilities inventory that

is capable of indicating the statewide percentage of facility utilization and projecting school facility needs five years in advance, in order to permit the board to study alternative proposals for the allocation of funds for new construction, maintenance, and rehabilitation.¹²⁵

Chart 11a reproduces a summary of the 2001 school accountability report card for Coronado Elementary School, one of the schools attended by named plaintiffs.¹²⁶ This report card indicates that Coronado Elementary "has gained 162 API points over a two-year period, ranking us 24th in the state of California and number one in the Oakland-San Francisco Bay Area." In addition, the report card indicates that:

- about a third of the school's students are ELLs,
- almost all students are SES_D,
- student/ teacher ratios are just below county and state averages,
- the percent of teachers with emergency credentials is nearly double the state average,
- the achievement of ELLs substantially exceeds county and state norms in reading, language and math,
- reading and math texts are dated 1999 or later with enough copies for each student,
- the number of computers per student is slightly above the state average,
- the 92% student attendance rate is just below the district average, and
- API awards were received for two consecutive years.

¹²⁵ CEC § 33126.1, § 33126.2, & § 33126.5.

¹²⁶ The full school accountability report card for this school can be found at www.cde.ca.gov/sarc/.

¹²⁴ CEC § 33126.1.

Summary input information, including ethnicity, parent education level, free/reduced lunch participation, English language learners, multi-track school, school mobility, district mobility, average class size, fully credentialed teachers, and teachers with emergency credentials, is also reported with each school's API score posted on the state website.¹²⁷ Chart 11b reproduces the 2002 API growth report for Coronado Elementary School. This report indicates that the school:

- substantially exceeded its schoolwide and comparable improvement growth targets and is eligible for awards (also met student participation and parent waiver requirements),
- enrolls 52% African-American and 44% Hispanic students,
- has increased the percent of ELLs to 39% and SES_D to 99%,
- is *not* a multitrack year-round school,
- has reduced the percent of teachers with emergency credentials to the state average,
- has kept average class sizes slightly below state averages, and
- had no students excused from testing by parent written request.

Intervention Process Uses Inputs.
Russell asserts:

By requiring schools to actively describe the impacts their inputs have on outputs, identify potential problem areas, and establish short and long term goals, educational benefits of accountability could be more fully realized.¹²⁸

¹²⁷ See www.cde.ca.gov/api.

¹²⁸ Russell Report, *supra* note 108 at xix.

Given that inputs affect outcomes and that at times it is the inputs that must be altered before outcomes are impacted, schools must be allowed and encouraged to set goals that focus first on the inputs.¹²⁹

But again, the school evaluation process described by Russell above already is part of California's accountability program. It is included in the intervention process mandated by the *Accountability Act*.

Intervention site teams and external evaluators are required to consider inputs when constructing and implementing their action plans for improving school performance. The use of multiple measures, active reflection, and goal setting by the school community, characteristic of many of Russell's projects,¹³⁰ is part of the design of the comprehensive, self-study process prescribed by the California legislature.¹³¹

Moreover, also similar to Russell's projects, the California system provides maximum flexibility for schools to select and prioritize improvement goals from a myriad of potentially relevant factors unique to local site conditions and needs. In the event that direct state intervention becomes necessary, the *Superintendent* is given broad powers to choose among a variety of options, including input options such as the reassignment of administrators and teachers.¹³²

¹²⁹ *Id.* at 57.

¹³⁰ See *Id.* at ii, iii.

¹³¹ See CEC § 52053 - § 52055.5.

¹³² *Id.*

At one point in his report, Russell acknowledges that California already does what he claims it should do:

This evaluation process is the closest California's accountability system comes to requiring schools to provide an account of their practices. Implied in this evaluation process is a desire for schools to take corrective action to improve problematic practices (whether they be curriculum misalignment, instructional practices, resource allocation, quality of teachers, quality of facilities and related educational materials, leadership, etc.).¹³³

Other groups have also followed California's lead in focusing on student outcome measures rather than inputs. In his report, Russell notes that the Western Association of Schools and Colleges, a private, nonprofit group responsible for accrediting public schools, has recently changed its evaluation criteria from a process model focused on inputs to an outputs model focused on assessing student achievement.¹³⁴

Russell's complaint seems to be that California should spend more money on the *Intervention Program* and make it mandatory for all schools. However, this suggestion is premature. As indicated earlier, AIR was selected as the external evaluator for the *Accountability Act*. The preliminary Phase I Report indicated positive results for the *Intervention Program* but much data remained to be collected. The final report will provide a more comprehensive evaluation of the results of the accountability program and recommendations

¹³³ Russell Report, *supra* note 108 at ix.

¹³⁴ *Id.* at 14.

for improvement.¹³⁵ At that time, the *Board* will have much more relevant information on which to base policy decisions about the API and the accountability program.

Evidence From an Unsuccessful, Court-imposed Inputs Experiment

In his report, Russell asserts:

Although student test scores have become the predominant form of "educational accountability" in most states, it is a seriously flawed approach to helping schools improve teaching and learning. . . . To improve the current [accountability] system, the types of information considered by the system must be expanded to include inputs . . . [O]ne first step toward improving the performance of students is to replace emergency credentialed teachers with teachers that are fully credentialed. . . . But teacher quality is only one of many inputs that may be in need of improvement. Others include adequate textbooks, curricular materials, access to current technology, classrooms and schools that are not overcrowded, sanitary conditions, and environment conducive to learning, etc.¹³⁶

However, Russell's proposals were attempted in Kansas City and they failed miserably.

In 1985, as a remedy for decaying school facilities, poor teachers, outdated instructional materials, and low-performing students, a federal judge imposed an inputs model similar to that described by Russell on the Kansas City, Missouri public schools. This fully-funded, inputs-

¹³⁵ AIR Evaluation Report, *supra* note 78.

¹³⁶ Russell Report, *supra* note 108 at 42, 45, 47.

based, educational experiment continued for more than a decade and produced dismal results.¹³⁷ The following excerpts from a 1998 report by Los Angeles education writer, Paul Ciotti, summarize the Kansas City experiment.

In 1985 a federal judge took partial control over the troubled Kansas City, Missouri, School District [] on the grounds that it was an unconstitutionally segregated district with dilapidated facilities and students who performed poorly. ... [He] ordered the state and district to spend nearly \$2 billion over the next 12 years [75% from the state; 25% from increased local taxes] to build new schools, integrate classrooms, and bring student test scores up to national norms. ... The idea was that Kansas City would be a demonstration project in which the best and most modern educational thinking would for once be combined with the judicial will and the financial resources to do the job right. No longer would children go to schools with broken toilets, leaky roofs, tattered books, and inadequate curricula.

... Kansas City spent as much as \$11,700 per pupil – more money per pupil, on a cost of living adjusted basis, than any other of the 280 largest districts in the country. The money bought higher teachers' salaries, 15 new schools, and [renovations in 54 others]. ... Included were nearly five dozen magnet schools, which concentrated on such things as computer science, foreign languages, environmental science, and classical Greek athletics. ... The ratio of students to instructional staff was 12 or 13 to 1, the lowest of any major school district in the country. ... For working parents the district provided all-day kindergarten for youngsters and before- and af-

ter-school programs for older students.

... Kansas City did all the things that educators had always said needed to be done to increase student achievement – it reduced class size, decreased teacher workload, increased teacher pay, and dramatically expanded spending per pupil – but none of it worked. Test scores stayed put, the three-grade-level achievement gap between blacks and whites did not change, and the dropout rate went up, not down.

... Eric Hanushek, a University of Rochester economist who testified as a witness regarding the relationship between funding and achievement [] in January 1997, [found that] the real problem in American public education wasn't so much financial as structural. There were no incentives in the current system to improve student performance – nothing rested on whether students achieved or not.

... **Postscript:** Confirmation from Sausalito (1998) People who believe there's a strong connection between money spent on education and student achievement have a hard time explaining what's going on in the tiny 284-student Sausalito, California, Elementary School District. The district spends more than \$12,300 per student each year – nearly three times the state average. Students go to school in freshly painted buildings, with manicured lawns and new playground equipment. Class size is a mere 16 students per room, half that of many larger districts. The district has special instructors for art, drama, science, and computers. Yet, when it comes to student achievement, none of that seems to matter. Test scores are the lowest in Marin County¹³⁸

¹³⁷ Ciotti, P., *America's Most Costly Educational Failure*, Investors Business Daily, April 29, 1998.

¹³⁸ Ciotti, P., *Money and School Performance: Lessons from the Kansas City Desegregation Experiment*, Cato Policy Analysis No. 298, March 16, 1998; Ciotti, P., *supra* note 137. 1998 Stanford Test, 2000 API and 2002 API data for Sausalito Elementary District compared to Marin County confirm the report's conclusion.

Validity of Test Data

Russell repeatedly criticizes the role of nationally-normed, achievement tests in the API. He asserts:

National, norm-referenced tests provide no information about student performance in specific subject areas.¹³⁹

In short, the [*Stanford Test*] is a poor instrument for either identifying student weaknesses within specific sub-domains or determining whether students have achieved acceptable levels of skills or knowledge within a given domain. And because California's accountability system is heavily dependent on the [*Stanford Test*], the system has little promise for helping schools identify strengths and weaknesses in student skill and understanding in specific areas of mathematics, language arts, social studies or science.¹⁴⁰

These criticisms are *not* valid. It is *not* the purpose of a standardized achievement test (e.g., the *Stanford Test*) to provide specific, diagnostic information for each student. Such tests are not long enough to do so. Neither are the California Standards Tests. A diagnostic test for a content area (e.g., mathematics) would need several items per individual skill and require days of total testing time to cover all the important skills. The cost of such testing on a state level

would be prohibitive and schools would reassert their recurring complaints of too much testing time.

The purpose of a standardized achievement test is to sample a domain of academic content knowledge and skills and to provide a global evaluation of student performance. When a student performs poorly on such a test, it is the responsibility of the school and the student's teacher to collect additional information to diagnose the reason(s).

Nonetheless, standardized achievement tests provide educators with clues about where to search for student academic deficits. The *Stanford Test* and the Standards Tests provide subscores that can be used to evaluate the relative strengths and weaknesses of an individual student within a content area. A sample student score report for the *Stanford Test* and Standards Tests is presented in Chart 12a and illustrates the type of subscore information reported.

This subscore information can be used to identify students' relative strengths and weaknesses. For example, on a mathematics test, a 7th grade student may correctly answer 90% of the number sense items, 75% of the algebraic concepts items, 80% of the geometry/measurement items, and 40% of the probability and statistics items. This hypothetical student has demonstrated a relative strength in number sense and a weakness in probability and statistics.

Stanford Test	1998 - NPR FOR AVERAGE STUDENT				
	2	3	4	5	6
	67	69	74	73	73
	29	44	--	50	44
	70	69	70	76	77
	57	70	30	41	33
	2000			2002	
	API	Rank	Cmpr to	API	Rank
	585	36/39	Marin	695*	31/41
	561*	37/39	Co.	625*	37/41
			Elms	499*	41/41

¹³⁹ Russell Report, *supra* note 108 at xiii.

¹⁴⁰ *Id.* at xv.

Further inquiry by the teacher might show that the student was absent a significant number of days during the probability and statistics unit and that the student's unit test performance indicated a clear lack of mastery of several important concepts. Those concepts could then form the basis for remediation activities for this student.

Russell acknowledges the importance of designing and using a test based on its purpose. He states:

The purpose of the test informs the type, quantity, difficulty, and, sometimes, order of items that form the test. While it is common practice to use a given test for purposes other than its intended use, this is not good practice.¹⁴¹

Russell reaffirmed this opinion in his deposition.¹⁴²

Testing CA Content Standards

Russell acknowledges:

The state has specified that tests employed as part of the accountability system should measure skills and knowledge specified in the curriculum frameworks from which schools are expected to teach.¹⁴³

However, Russell also asserts:

The [*Stanford Test*] ... is not aligned with California standards.¹⁴⁴

¹⁴¹ *Id.* at xiv.

¹⁴² Russell Deposition, *supra* note 121 at p. 379, line 17 - p. 380, line 14.

¹⁴³ Russell Report, *supra* note 108 at x.

¹⁴⁴ *Id.* at 8.

Teachers must choose whether to focus instruction on the skills and knowledge emphasized in the standards or on the misaligned content of the [*Stanford Test*].¹⁴⁵

This is not true.¹⁴⁶ In fact, there was significant overlap between the *Stanford Test* and the California content standards (see Table 1 and earlier section on *Augmentation of the Stanford Test*) and the *Advisory Committee* clearly indicated that use of only the *Stanford Test* scores was an interim decision until the Standards Tests could be developed and implemented.

In addition, the *Stanford Test* included enabling skills from prior grades that are important prerequisites for the more challenging content in the California standards. Thus, when students score poorly on the *Stanford Test*, schools are justified in teaching the enabling and lower level content skills it measures as a bridge to the more challenging skills included in the California content standards. It is also considered good teaching practice to review the work of the prior grade at the start of a new school year and to review prerequisite skills before beginning a new content standard.

Moreover, the augmented tests provided early estimates of student mastery of all of the California con-

¹⁴⁵ *Id.* at x-xi.

¹⁴⁶ Although the basis for this erroneous conclusion is not clear, Russell may have been relying on an incorrect statement in the 1999 final report of the *Advisory Committee*. CDE, *The 1999 Base Year API: The Report of the Advisory Committee for the PSAA*, Oct. 1999 (The norm-referenced component of this test is not linked to California content and performance standards), p. 2.

tent standards at a grade level. These augmented test scores included a substantial number of *Stanford Test* items at each grade level that the *Department* judged to match the California content standards plus additional items written to measure portions of the California standards not covered by the *Stanford Test*.

Table 2c (bottom) summarizes when *Stanford Tests* and Standards Tests in reading and mathematics were given, when they were included in the API and what weights were used across the years 1999-2002. The top of Table 2c summarizes the expected difference in the statewide 2002 base API results had it been calculated based only on *Stanford Test* results versus only on Standards Tests results. Using both types of tests with the normal weights, the state API was 661.5. Using *Stanford* only it was higher; using Standards Tests only it was lower.

These data demonstrate that on average schools would have scored worse had their APIs been based only on the Standards Tests because these tests are more difficult – note the lower percent of students in the top two performance levels. For example, in math, 50% of California students scored in the top two performance levels on the *Stanford Test* but only 33% scored in the top two performance levels on the Math Standards Test. The data in Table 2c demonstrate that the phase-in of the Standards Tests was a reasonable compromise for establishing performance on enabling skills first and then moving schools toward the

more demanding standards-based content when the Standards Tests were fully developed.

In the end, Russell seems to concede that, while not perfect, California's use of the *Stanford Test* was reasonable under the circumstances. As Russell acknowledges:

Clearly, this decision was a matter of judgment and, without a better-aligned test in hand, was deemed the best alternative.¹⁴⁷

Replacing the *Stanford Test* with the *California Test*

Russell suggests that California's replacement of the *Stanford Test* with the *California Test* will not improve its accountability system. He asserts:

[T]he expected change from the [*Stanford Test*] to a new [norm-referenced test] in 2003 does not rectify the issue of poor alignment. Like the [*Stanford Test*], the new test will be a general test of skills and knowledge that was designed to provide normative comparisons at the national level. Moreover, like the [*Stanford Test*], the new [norm-referenced test] will not be developed to specifically target skills and knowledge specified by California's standards.¹⁴⁸

[T]he introduction of the [*California Test*] to replace the [*Stanford Test*] in 2003 may also disrupt the comparability of that year's API with previous years.¹⁴⁹

Like the [*Stanford Test*], the [*California Test*] is a nationally norm-

¹⁴⁷ Russell Report, *supra* note 108 at 30.

¹⁴⁸ *Id.* at xi.

¹⁴⁹ *Id.* at 12.

referenced test that provides poor diagnostic information at the student level.¹⁵⁰

But again, Russell's criticisms are unfounded. The *California Test* is a valid and reliable, nationally-normed, standardized achievement test designed to measure similar content skills and for the same purpose as the *Stanford Test*. When introduced in 2003, it will count for only 20% of the API at the elementary and middle school levels and only 12% of the API at the high school level. The remainder of the API weight, 80% for elementary/middle school and 88% for high school, will be based on Standards Tests specifically written to measure the California content standards.

Thus, the 2003 base API will consist primarily of Standards Tests. In future years, as additional Standards Tests are added in science and social science, the API weight of the *California Test* will drop still further.

In addition, contrary to Russell's claims, API comparisons will *not* be disrupted because scores on the *California Test* will be linked statistically to the *Stanford Test*. This will provide an appropriate 2002 base API and 2003 growth API comparison. The 2003 base API will then be computed using actual norms for a valid 2003-04 reporting cycle.

Plans for Additional Non-Test Indicators

Russell acknowledges that:

Test scores are not the only components under consideration for inclusion in the accountability index. The [Accountability Act] legislation mandates that measures such as student and teacher attendance rates and high school graduation rates be incorporated into the API calculation. Additional measures ... "are to be included only when available, valid, and reliable."¹⁵¹

As indicated earlier in the section *Evolution of the API*, the feasibility of adding the attendance rate and graduation rate indicators listed in the *Accountability Act* to the API is being studied by the *Department*. However, the statute is clear that lawmakers envisioned incorporating such measures only if they are available, valid and reliable. Russell makes no claim that these criteria have been met.

As indicated previously, there are currently problems with attendance and graduation data with respect to all three statutory criteria. Thus, it is not reasonable to incorporate these measures in the API at this time. Nonetheless, the *Department* and the *Board* are continuing to study alternatives and to work toward a statewide student information database.

NAEP Comparisons

Despite criticizing California for relying in part on a nationally standardized achievement test partially aligned to California content standards, Russell criticizes the California accountability system for demon-

¹⁵⁰ *Id.* at xv.

¹⁵¹ *Id.* at 11, quoting CEC § 520252(b).

strating improvement that is not totally consistent with gains on another nationally standardized test not fully aligned with California standards: the National Assessment of Educational Progress (NAEP or *National Assessment*). Russell asserts:

the sharp increases in California on the [*Stanford Test*] do not generalize to the NAEP.¹⁵²

NAEP has a different purpose than the *Stanford Test* and the California Standards Tests. NAEP is designed to provide a periodic snapshot of achievement by a national sample of students in the U.S. in selected content areas and grades. Unlike the California tests, NAEP is not administered annually in all grades, does not produce individual student scores, does not report student scores to parent or schools, does not measure longitudinal growth of individual student or schools, and is not used to allocate rewards or remedial resources to schools or districts.

In contrast to the California test for which all students can be compared on the same set of test items, NAEP uses a matrix sampling approach in which no student takes a full length NAEP test and there are no individual student scores. Thus, it is not possible to determine whether a given student is proficient in a content area or how many students in a school are proficient in a content area (e.g., reading, math). Because students and schools do not receive results, students are not likely to be highly motivated when taking the

NAEP test and there is little incentive for schools to be concerned about their performance.

NAEP used to be administered every four years but has recently begun a two-year testing cycle in which reading and math tests are administered every other year in grades 4, 8 and 12. This contrasts with California's annual administration in grades 2-11. Although both the California Standards Tests and NAEP use the same performance descriptors to classify students (e.g., basic, proficient), these terms are defined differently. Different standard setting methodologies on different content standards were used to determine the level of student performance required for each classification on the two different tests.

In addition, because the California Standards Tests are based on the California content standards for which schools are held accountable, one would expect instruction focused on state standards to have a greater impact on student performance on the California Standards Tests than on NAEP tests. That is, it would not be unreasonable for scores on the state test to increase faster on average than scores on the NAEP tests. Nonetheless, state determinations of proficiency for a grade and content area should not be extremely different from NAEP results. For example, if a state test indicated that 80% of its students were proficient in math and NAEP indicated 25% proficient, there would be cause for concern.

¹⁵² *Id.* at 27.

Keeping in mind the differences and limitations stated above, and using the Standards Tests with similar performance standard categories rather than the *Stanford Test*, trends in California and U.S. NAEP data and California state test data can be examined. Table 5 and Chart 17 summarize these data by ethnicity for reading and mathematics in grades 4 and 8. Note that the latest available data on NAEP proficiency is for 2000 while the earliest California Standards Tests proficiencies are for 2001.

Up to 1999-2000 when California standards testing began and the first APIs were calculated, NAEP reading scores increased slightly in grade 4 and somewhat more in mathematics in both grades. However, the percent of proficient students remained low and lagged behind U.S. performance.

Data in Table 10 and reported by Achieve comparing percent proficient for NAEP and state standards tests demonstrate that "states have had different levels of achievement in mind when defining proficiency."¹⁵³ However, these data also indicate a wide variability with many states demonstrating a much larger discrepancy than California.

The results from the Standards Tests, which are based on the NAEP model, suggest that students may improve their performance in 2002 and 2004. Nonetheless, for all the reasons described above, one would expect performance on the state test to be somewhat higher than on

¹⁵³ Achieve, Inc., *Staying on Course*, 2002, p. 11.

NAEP. As long as state standards are challenging (as in California) and as long as NAEP and state results are not so far apart as to be unreasonable, state results are meaningful indicators of student proficiency on state standards.¹⁵⁴

Policy Decisions

The *API Framework* adopted by the Board specifically articulated a goal of fairness to all schools in the development of the API. It stated:

Given differences in student backgrounds and the resources available to schools, not all schools and students start out the same. . . . The challenge is to reflect differences among schools and students fairly *without institutionalizing lower expectations for some*. . . . [The API] should [] be constructed in such a way that improvement is possible regardless of current level of performance (i.e., schools with high-achieving students and schools with low-achieving students can both receive rewards).¹⁵⁵

Table 6a summarizes statewide API school performance and Table 6b and Chart 18 present a comparison across reporting cycles of the awards eligibility of California schools by state decile rank for the base API of each reporting cycle. Table 6b also includes a summary of reasons why some schools did *not* receive 2002 API growth results.

The data in Table 6a demonstrate several trends across the three API reporting cycles. The percent of

¹⁵⁴ See discussion in section on *Lessons From Other States* and Table 10, *infra* p. 69, for comparisons with larger discrepancies in several other selected states.

¹⁵⁵ API Framework, *supra* note 32 at 2-3, emphasis added.

schools meeting their growth targets decreased from 71% to 53%. However, during this period, more stringent participation and parent waiver requirements were adopted, the *Language Arts Standards Test* was added to the API, and minimum growth requirements were increased.

Also across the three API reporting cycles, the percent of schools with an increased schoolwide API decreased from 89% to 69%, the percent of schools with APIs at or above 800 increased from 12% to 20%, median API scores increased from 620 in 1999 to 647 in 2002, and the percent of schools eligible for awards decreased from 67% to 39%. Schools are improving but it is becoming harder to meet all the criteria for awards eligibility.

The data in Table 6b and Chart 18 summarize awards eligibility by decile. In the first reporting cycle, awards eligibility ranged from the low 60s in the lower deciles to the low 70s in the upper deciles, a reasonably equitable distribution that allowed many low-performing schools to qualify for awards. However, with the introduction of more stringent minimum growth requirements in the upper deciles, the data for the latest API cycle have reversed the trend with the percent of awards eligible schools in the 50s in the lowest deciles dropping to the low 30s in the highest deciles.

As the data in the bottom of Table 6b indicate, about a quarter of California schools did not receive APIs in 2002. Over half were alternative or very small schools. Only 2% were due to

testing irregularities and 4% to excessive parent waivers.

General Policy Disagreements

Russell asserts a variety of general disagreements with California policymakers' choice of the API to summarize school performance. These general disagreements tend to be incorrect or misleading statements about characteristics or presumed effects of the API. Examples of particular statements (S) followed by a correction (C) are presented below.

S An API of 800 (or any value for that matter) does a poor job of characterizing the actual performance of students in a school.¹⁵⁶

C The API is a summary statistic designed for a summative evaluation of school performance. When that statistic indicates problems, a more detailed analysis is called for. Schools receive much more detailed test information in the *Assessment Act* reports. An example is presented in Chart 12b.

S Unless API score increases are above average, they go unnoticed.¹⁵⁷

C Schools are not required to increase their APIs more than the average for schools in the state to be eligible for awards. Schools are required to gain 5% of the difference between their previous year APIs and the interim target of 800 (or at least five score points). That growth target may be more or less than the state average school growth. As indicated in Table 6b,

¹⁵⁶ Russell Report, *supra* note 108 at 21.

¹⁵⁷ *Id.* at xv.

more than half the schools in the lowest 5 deciles (below the state average) were eligible for awards in 2002.

S Although often misinterpreted as showing no growth, percentile ranks that remain the same across years actually represent substantial growth – growth that is identical to the average student nationwide.¹⁵⁸

C Students must learn some new content to maintain their percentile ranks from one year to the next. However, this does not represent growth identical to the average (50th percentile rank student) but rather the growth that is required at the students' obtained percentile ranks. That is, the amount of individual student growth necessary to maintain a percentile rank varies depending on the percentile rank. Typically that growth is greater for above average students than for below-average students.

Therefore, for students who are low-performing (e.g., 2 grade levels behind in reading skills), they will not catch up to their peers by maintaining their percentile ranks – they must gain more. Thus, the API gives schools relatively more credit for moving such students out of the lowest performance bands.

For example, on the *Stanford Test*, schools are awarded 300 more points for moving students from percentile ranks in the teens to percentile ranks in the 20s and 200 more points for moving students from percentile ranks in the 30s to percentile ranks in the 40s. Similar incentives

¹⁵⁸ *Id.* at xv.

on the California Standards Tests reward schools for moving students from the *far below basic* to the *below basic* performance category and from the *below basic* to the *basic* performance category. These incentives encourage schools to target their resources at the most educationally needy students.

S Tests that provide student-level data provide poor school-level data. A single test administered to all students within a school, whether it be norm-referenced, like the [*Stanford Test*] and [*California Test*], or criterion-referenced, like the High School Exit Exam, is inadequate for diagnosing instructional strengths and weaknesses within the school (or individual classrooms). ... Matrix sampling is a far more efficient and informative approach to collecting diagnostic information that can be used by teachers and schools to improve curriculum and instruction.¹⁵⁹

C As already indicated in the section *Validity of Test Data*, the California tests are *not* designed to be diagnostic and it would *not* be cost effective for them to do so. Schools should be able to use the test information they receive as clues for further investigation of group and individual student strengths and weaknesses (see Charts 12a and 12b).

S Although matrix designs do not provide comparable student-level scores, California's current accountability system only requires such student-level scores for the High School Exit Exam. For all other grade levels, a matrix design would be far more informative than is the current practice of administering the same set of test items to all students in a school and across the state.¹⁶⁰

¹⁵⁹ *Id.* at xvi.

¹⁶⁰ *Id.* at xvi.

C The *Assessment Act* specifies comparable student level scores to provide normative information to parents. It states:

It is the intent of the Legislature [] to provide a system of individual assessment of pupils In designating an achievement test, [the Board shall consider]: [the a]bility of the publisher to produce valid, reliable individual pupil scores.¹⁶¹

In addition, to be fair to schools, they should be held accountable for the performance of all students on a common set of skills, and students should be motivated when taking the test. Students are more motivated when they receive individual scores and interpretive information (e.g., percentile ranks, performance scores, subarea scores) that would not be available with a matrix sampled test.

S Once targeted, a school becomes eligible for funding that supports an investigation into conditions that may be negatively impacting student performance. The schools are then expected to remedy these conditions, but the extent to which the conditions are actually remedied is never examined.¹⁶²

C School site teams, the external evaluator, and the school district are required by law to monitor an intervention school's progress in implementing its action plan.¹⁶³ These schools also receive annual feedback on their efforts from the state through their state test score reports, API scores, decile rankings and awards eligibility. In addition, these schools are required to revise

¹⁶¹ CEC § 60602(a), § 60644(a).

¹⁶² Russell Report, *supra* note 108 at 59.

¹⁶³ CEC § 52053 et seq.

their school report cards each year to reflect changes in input and output measures.

S A standard setting process known as *item mapping* was used to define the four performance bands [on the Connecticut standards test], with an acknowledgement that the standards are set at a high level. . . . However, it is not clear whether any systematic approach involving expert judgment or based on test items was used [in California].¹⁶⁴

C Performance standards were set on the California standards tests using the *item mapping* procedure Russell cites as desirable. The *Stanford Test* quintiles used in the API calculations were selected to correspond to the five performance levels to be set for the Standards Tests.

Alleged Unfairness of API Policies

Several of Russell's criticisms of API policies are contradictory and self-defeating. For example:

➤ Russell asserts that the interim target of 800 is too high since differences in CA demographics (e.g., 20% English Language Learners versus 2% nationally) destine too many schools for failure. But he also asserts that an annual 5% growth rate toward the 800 goal is too little because it will take too many years for schools to attain the goal.¹⁶⁵

➤ Russell asserts that schools get new students each year from feeder schools, immigration and

¹⁶⁴ Russell Report, *supra* note 108 at 64.

¹⁶⁵ *Id.* at 21 & endnote xvi.

general mobility. But he acknowledges that student scores are not included in the API until a student has been in the school for at least a year.¹⁶⁶

➤ Russell asserts that schools have to work harder with disadvantaged populations to achieve growth targets. But the Accountability Act states that the Board “may set higher growth targets for the lowest performing schools because they have the greatest room for improvement” and its comparable improvement requirement for awards is intended to encourage schools to work harder with these challenging subgroups.¹⁶⁷

➤ Russell asserts that students must take harder tests each year as grade level increases. But he notes that a universal increase of 2 PRs – about 1 more question correct per student on the *Stanford Test* – would produce satisfactory annual growth for a school with an API of 480). Moreover, student growth for the API is not measured longitudinally from one grade to the next but by successive annual cohorts of students in the same grades.¹⁶⁸

➤ Russell asserts that the API does not diagnose or explain patterns of individual students’ learning. But this is not the stated purpose of the API, and *Stanford Test* content cluster scores and

Standards Tests subscores provide a starting point for student and school diagnosis.¹⁶⁹

Second-Guessing Policymakers

Russell faults the API for a variety of reasons that imply that policymakers should have made different decisions. For example, Russell asserts:

Generating an API from [*Stanford Test*] test scores requires an arcane calculation process.¹⁷⁰

California’s accountability system is a product of questionable policy decisions made by state officials.¹⁷¹

Although the end goal of the [API] is to summarize school performance with a single, seemingly precise number, the factors and weightings used to produce that single number are based on informed, but nonetheless subjective decisions. While decisions about some of these variables were informed by simulations and modeling conducted by members of [the *Tech-Group*], it is not clear how scientific the decision-making process was.¹⁷²

He further asserts that “alternative decisions were possible” and provides examples of “how minor changes in those decisions can have major effects.”¹⁷³

Reasonable, Judgmental Process. Constructing an academic performance indicator for schools is a judgmental, not a scientific process and

¹⁶⁶ *Id.* at 22.

¹⁶⁷ *Id.*; CEC § 52052(c).

¹⁶⁸ Russell Report, *supra* note 108 at 23, endnote xv.

¹⁶⁹ *Id.* at 24; see sections of this report on *Stated Intent of the Accountability Act and Validity of Test Data*, *infra*, p. 29 & p. 37. See also sample reports in Charts 12a & 12b.

¹⁷⁰ Russell Report, *supra* note 108 at 8.

¹⁷¹ *Id.* at 28.

¹⁷² *Id.*

¹⁷³ *Id.* at 29.

there is no gold standard that specifies what decisions should be made. The outcome of this judgmental process is defensible if the entity with the appropriate authority to do so makes the decision based on relevant advice, data and information from informed sources and stakeholders.

In California, decisions about the calculation of the API were made by the *Board* with advice from the *Advisory Committee*, the *Tech Group*, the *Superintendent* and the *Department*. The *Accountability Act* specifically delegated authority to the *Board* to make these decisions.¹⁷⁴ The *Accountability Act* also directed the *Superintendent* to appoint a representative and diverse *Advisory Committee* staffed by the *Department* to advise the *Superintendent* and the *Board* about all matters related to the creation of the API, awards and the intervention program.¹⁷⁵

The *Advisory Committee* and the *Tech Group* based their written recommendations on extensive meetings and debate among their members. Representatives of these groups also attended *Board* meetings to make presentations on the issues and to answer questions. In addition, the *Board* held public hearings on the API Regulations at which stakeholders had an opportunity to comment.

The written papers, presentations, agenda items, *Board* minutes and *Advisory Committee* minutes that document the decision-making proc-

ess clearly articulate a rationale for each API decision. They also indicate that the *Board* and its advisors collected relevant data, considered alternatives, and balanced competing policy goals.

Calculation Rationales. At the time the *Board* adopted procedures for calculating the API, the *Advisory Committee*, with assistance from the *Tech Group*, provided recommendations for each required decision along with a written policy rationale. They included:

- Choice of the **200-1000 scale** as simple to communicate but avoiding confusion with percentiles and percentages (0-100 scale) and providing a nonzero minimum score because even low-performing students have attained some level of academic proficiency;
- National percentile ranks divided into five bands** because the purpose of the nationally normed test is to obtain national comparisons, fewer levels are preferable, five levels are compatible with performance standards for the *Standards Tests* to be added in the future, and data simulations indicated no advantage in sensitivity to school status and growth measures for 10 performance bands.
- Progressive weights for the five performance bands** to encourage schools to focus on low-performing students by giving them more credit for

¹⁷⁴ CEC § 52050 et seq.

¹⁷⁵ CEC § 52052.5.

moving students out of the lowest performance bands and because data simulations indicated no disadvantage to high-performing schools in demonstrating growth.

- **Content area weights that reflect curricular priorities** (emphasis on reading in elementary grades and equal time per subject in high school) and are the same as those adopted earlier for identifying the first cohort of schools eligible for the *Intervention Program*.
- An **interim statewide performance target of 800** representing exemplary performance to which all schools should aspire. Based on 1999 *Stanford Test* data, the Technical Design Group estimated that only a very small percentage of schools attained it (8% elementary, 6% middle, and 4% high schools) and all other schools could earn awards by demonstrating 5% growth (specified statutory minimum) toward it.
- A **comparable improvement standard** for numerically significant ethnic and SES₀ subgroups of **80% of the overall school growth target** to compensate for the stringency of a conjunctive model and defining a subgroup of 100 or more as numerically significant (even if not 15% of the population) because that is the threshold

for number of students required to report an overall school API growth target.

- Use of standard statistical procedures to generate a **school characteristic index** (predicted API) using the list of characteristics specified in the statute for purposes of reporting similar schools API growth and ranks.¹⁷⁶

Rationales for each recommendation considered the principles of the API framework and the goals of the API. For example, performance bands were chosen over averages for weighting student performance for a school because "this methodology values gains by pupils at both the high and low end of the distribution of scores" and "best responds to the intent of the [*Accountability Act*]."¹⁷⁷

In addition, regarding weights, the *API Framework* stated:

In order to arrive at a summary statistic for an assessment, weights must be assigned to each content area. Ultimately, the value of these weights is a policy question. The weight that is assigned to a content area is an expression of the relative importance that the *Superintendent* and the *Board* attaches to that content area.¹⁷⁸

The only requirement in the *Accountability Act* regarding weighting was that test indicators must constitute at least 60% of the API.¹⁷⁹ The advisors recommended and the

¹⁷⁶ The 1999 Base Year API, *supra* note 35 at 6-20.

¹⁷⁷ API Framework, *supra* note 32 at 7.

¹⁷⁸ *Id.* at 7-8.

¹⁷⁹ CEC § 52052(a).

Board approved content area weights that reflected relative instructional time and importance.¹⁸⁰

Balancing Policy Tradeoffs. In the *API Framework* adopted by the *Board*, the *Advisory Committee* further stated:

Decisions in developing the API will involve trade-offs between technical soundness and efficiency, but fairness must not be sacrificed. The challenge will be to balance these issues while also considering legal requirements, data availability, and sound education policy.¹⁸¹

Extensive debates among advisors, Board Members and stakeholders occurred prior to *Board* adoption of final API regulations. The legislature ratified many of these policy decisions by amending the *Accountability Act* in 2000 with SB 1552.

Reconsideration of API Policies. In September 2001, the *Board* considered recommendations from the *Tech Group*, the *Advisory Committee*, the *Superintendent*, and the *Department* regarding procedures to be followed in incorporating the *Language Arts Standards Test* into the API. At that time, the *Board* was given the opportunity to reconsider previous API policy decisions and to decide which of several proposed API calculation options should be selected for adding the *Language Arts Standards Test* to the API.

To assist the *Board* in examining and evaluating alternatives, the *Tech*

¹⁸⁰ SBE, Minutes & Agenda Item #23, Nov. 9, 1999, Attachment 1: Advisory Committee, *The 1999 Base Year API*.

¹⁸¹ API Framework, *supra* note 32 at 2.

Group prepared a policy paper that identified the following 6 issues in need of resolution, proposed options for addressing each issue, and made recommendations.

1. Should the range of 200 to 1000 and the performance target of 800 be retained?
2. What are the performance level weighting factors that will be used to calculate the single number that will summarize pupil performance [on the *Language Arts Standards Test*]?
3. How will the summary number for [the *Language Arts Standards Test*] be integrated into the API?
4. What weight will be given to the [*Language Arts Standards Test*] relative to the [*Stanford Test Language Arts component*]?
5. Should this weight be applied immediately in 2001 or phased in gradually? and
6. When new components such as the [*Language Arts Standards Test*] are brought into the [API], should the average value of the API be maintained?¹⁸²

The *Board* adopted the recommendations of the *Tech Group* on these issues with one change regarding

¹⁸² SBE Minutes & Agenda Item # 17, Sept. 5, 2001, Attachment 1: Technical Design Group, *The 2001 Base API: Integrating the California Standards Test for English-Language Arts into the API*, July 18, 2001, p. 2.

weights proposed by the *Superintendent*. In concurring with the recommendations of the *Tech Group*, the *Advisory Committee* supported retention of as many features as possible from the API that had been in effect for 2 years.¹⁸³

In January 2003, in conjunction with deliberations finalizing the 2002 base API, the *Board* changed the component weights of the API to incorporate new Standards Tests and to significantly decrease the weight of the *Stanford Test*.¹⁸⁴ In addition to emphasizing the closely aligned Standards Tests, the decrease in weight of the *Stanford Test* was intended to maintain the stability of the API in the changeover from the *Stanford Test* to the *California Test*.

Russell Alternatives Considered and Rejected. In Appendix B, Russell presents several alternatives to the API calculation decisions made by the *Board*. Each of these alternatives was considered and rejected by the *Tech Group*, the *Advisory Committee*, the *Superintendent*, and the *Department* during the policy deliberations and decision-making process described above. These alternatives, together with the corresponding *Tech Group* issues from the 1999 and 2001 policy papers that considered and rejected them, are summarized below.¹⁸⁵

¹⁸³ SBE Minutes & Agenda Item # 17, Sept. 5, 2001, Attachment 2: Letter of Recommendations from the Co-chairs of the *Advisory Committee*, Aug. 21, 2001.

¹⁸⁴ SBE Minutes & Agenda Item #5, Jan. 8, 2003, Attachment 1: 2003, Attachment 1: Technical Design Group, *Revisions in the 2002 Base Academic Performance Index (API)*.

¹⁸⁵ Russell Report, *supra* note 108 at 68-74; SBE Minutes & Agenda Item #17, Sept. 5, 2001, Attachment 1, *supra* note 182 at 2-8; SBE Minutes & Agenda Item #

<u>Russell Alternative</u>	Corresponding 2001 & 1999 Tech Group Rejected Options
➤ Different Performance Bands	<input type="checkbox"/> Issue 2, #2 Issue 1G, #2
➤ Different Performance Band Weights	<input type="checkbox"/> Issue 2, #2, #3 Issue 1H, #2
➤ Different Content Area Weights	<input type="checkbox"/> Issue 4, #1 Issue 1I, #2
➤ Different Performance Target for Schools	<input type="checkbox"/> Issue 1, #2 Issue 2A, #2.

Summary. The actions by the *Board* in making and reaffirming policy decisions regarding calculation of the API were consistent with the guiding principles from the previously adopted *API Framework*, based on detailed study and data simulations by the *Tech Group*, recommended by the *Advisory Committee* following extensive debate, endorsed by the *Superintendent* and the *Department*, and consistent with the assessment and reporting requirements in the federal *No Child Left Behind Act*. In consideration of all of the foregoing information and reasons, it is my professional opinion that the process California used to create the API was appropriate and reasonable and therefore the decisions of the *Board* should be preserved.

Consistency with Federal Law

Recent federal legislation has mandated that states establish academic content standards and annual assessments in reading and mathematics for all students in grades 3-8 and at least one grade in high school by

23, Nov. 9, 1999, 1999 Base Year API *Advisory Committee Report*, at 8-11, 14.

2006.¹⁸⁶ Science assessments must be administered in at least one grade in elementary, middle and high school by 2006.¹⁸⁷ This federal law, known as the No Child Left Behind Act of 2001 (NCLB or *Federal Education Law*) and funding remedial education for low-performing, disadvantaged students, states:

The State shall have such **academic standards ... including at least mathematics, reading or language arts, and (beginning in the 2005-2006 school year) science**, which shall include the same knowledge, skills, and levels of achievement expected of all children.

Standards under this paragraph shall include –

(i) **challenging academic content standards in academic subjects** that –

- (I) specify what children are expected to know and be able to do;
- (II) contain coherent and rigorous content; and
- (III) encourage the teaching of advanced skills; and

(ii) **challenging student academic achievement standards** that –

- (I) are **aligned** with the State's academic content standards;
- (II) describe two levels of high achievement (**proficient and advanced**) that determine how well children are mastering the material in the State academic content standards; and
- (III) describe a third level of achievement (**basic**) to provide complete information about the progress of the lower-achieving children toward mastering the proficient and advanced levels of achievement.

...

Each State plan shall demonstrate that the State has developed and is implementing a **single, statewide State accountability system** that will

be effective in ensuring that all [schools] make adequate yearly progress Each State accountability system shall –

(i) be based on the academic standards and academic assessments adopted [under this act] and other academic indicators [consistent with this act] and shall take into account the achievement of all public [school] students;

(ii) be the same accountability system the State uses for all public [schools]; and

(iii) include sanctions and rewards, such as bonuses and recognition, the State will use to hold [schools] accountable for student achievement and for ensuring that they make adequate yearly progress in accordance with the State's definition

...

Adequate yearly progress shall be defined by the State in a manner that –

(i) applies the same high standards of academic achievement to all public [school] students in the state;

(ii) is **statistically valid and reliable**;

(iii) results in continuous and substantial academic improvement for all students;

(iv) measures the progress of public [schools, districts] and the State **based primarily on the academic assessments** [described in this act];

(v) includes separate measurable annual objectives for [growth] for [all students, economically disadvantaged students, major racial and ethnic groups, students with disabilities, and students with limited English proficiency]; except that disaggregation of data ... shall not be required [if] the number of students in a category is insufficient to yield statistically reliable information or the results would reveal personally identifiable information about an individual student;

(vi) ... includes **graduation rates ... and at least one other academic indicator ...** [that are] **valid and reliable**, and are consistent with relevant, nationally recognized professional and technical standards, [and] **may not use those indicators to reduce**

¹⁸⁶ NCLB, Title I, § 1111 (January 7, 2002), 20 U.S.C. § 6301 - § 6578.

¹⁸⁷ *Id.*

the number [of schools] that would otherwise be subject to school improvement, corrective action, or restructuring ... but may use them to identify additional schools

...
Each State plan shall contain assurances that –

...
the State will, beginning in [2002-2003], participate in biennial State academic assessments of 4th and 8th grade reading and mathematics under [NAEP] if the Secretary pays the costs of administering such assessments.¹⁸⁸

California has already adopted the academic content standards required by the *Federal Education Law*. The *Assessment Act* tests administered in grades 2-11 and the *High School Exit Exam* administered for the first time in grade 10 satisfy the testing requirements of the *Federal Education Law*. *Education Week* reported that at the time the law was signed, only 9 states, including California, administered the required mathematics and language arts tests aligned with their standards in grades 3-8.¹⁸⁹

California has proposed a two-part school accountability system – a combination of the current API and an additional percent proficient criteria – to satisfy the adequate yearly progress requirement of the *Federal Education Law*. Comparable to the *California Accountability Act*, the *Federal Education Law* requires school accountability to be based primarily on academic assessments and to include other indicators only when valid and reliable.

¹⁸⁸ *Id.*, emphasis added. See also, C.F.R. § 200 et seq.

¹⁸⁹ Quality Counts 2002, EDUCATION WEEK, January 9, 2002.

If school accountability based primarily on academic testing were to be eliminated in California in favor of the input variables suggested by the plaintiffs, California would no longer be in compliance with the *Federal Education Law* and could lose vital Title I federal funding. Since the schools receiving the most Title I funding tend also to be those with the highest percentage of *SES_D* and minority students, this potential loss of revenue would negatively impact those schools that most need it.

Aggregation Across Grades

Russell criticizes aggregation of scores at the school level:

Aggregating scores at the school level masks the successes and failures at the grade and classroom levels.¹⁹⁰

Yet he acknowledges that aggregation at the grade or classroom level presents problems of its own:

While aggregation at the grade or classroom level may be a poor fix for this problem, it might promote closer examination of practices and issues within these smaller operational units.¹⁹¹

Moreover, Russell ignores the fact that classroom data is available to schools that want it. Results from the *Stanford Test* and *Standards Tests* as prescribed by the *Assessment Act* provide data by grade level within school for all students and selected subgroups.¹⁹² In addition, in-

¹⁹⁰ Russell Report, *supra* note 108 at xii.

¹⁹¹ *Id.* at xiii.

¹⁹² See www.cde.ca.gov/star/.

dividual schools can order classroom summaries for the *Stanford Test* directly from the publisher or can calculate their own classroom summaries from their student test data files.

Error in the Accountability System

The *API Framework* adopted by the Board listed accuracy as one of its guiding principles, stating:

The API must measure school performance and growth as accurately as possible. Changes from the base year to the growth year within each overlapping two-year API cycle should reflect actual changes in school performance, not changes in testing procedures, inclusion criteria, or other variables.¹⁹³

Russell asserts that API scores have too much measurement and sampling error to be useful indicators of school performance. However, given reasonably reliable measures, measurement error has a negligible effect on the accuracy of observed changes in school averages over time. The important source of error that should be considered when making inferences about the quality of a school over time is sampling error.

Test Score Error

Russell asserts:

Measurement error impacts the reliability of scores and score changes, so individual test scores will always be to some degree volatile.¹⁹⁴

¹⁹³ API Framework, *supra* note 32 at 3, emphasis in original.

¹⁹⁴ Russell Report, *supra* note 108 at 23.

Individual students can have a bad day or a lucky day and have their achievement over or under estimated by a standardized achievement test. However, student error is closely related to test reliability. Nationally-normed achievement tests such as the *Stanford Test* with sub-test reliabilities in the high 80s and low 90s, have small standard errors of measurement.

Moreover, individual student error is not a factor in the API calculations because the school is the unit of analysis for the API. When student scores are combined to form a school statistic, the net effect of the random individual student positive and negative measurement errors should be approximately zero. Potential volatility of school APIs comes not from errors in measuring individual students but from differences in cohorts of students from one year to the next.

In addition, Rogosa has demonstrated that although year-to-year improvements in school APIs tend to be more variable for smaller schools than larger schools, they are unrelated to the magnitude of the standard error of the improvement.¹⁹⁵ That is, larger API growth is *not* associated with larger errors and vice versa. The scatterplots in Chart 13 illustrate these relationships. Rogosa has also demonstrated that the primary use of the API to measure schools' success in meeting their individual growth targets, as opposed to rank ordering schools, makes ac-

¹⁹⁵ Rogosa, D., *Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger "Volatility in School Test Scores"*, Stanford University, October 2002, Section 2.

curacy of classification a much more important indicator of error than estimates of sampling error (or their associated reliabilities).¹⁹⁶

API Error

Russell asserts:

[A]ggregate (or mean) test scores for schools containing fewer than 100 to 140 students fluctuate substantially from year to year. These fluctuations result largely from error in measurement and differences in the characteristics of cohorts rather than real differences in learning¹⁹⁷

[A]ggregate test score error was not fully openly disclosed by the State until July of 2002 and was reported to be approximately 20 points. This 20 point error means that the API score for an "average" school could be 20 points higher or 20 points lower than the actual score reported by the State. For many schools, test score error is as large as the amount of improvement prescribed by the State.¹⁹⁸

The *Accountability Act* specifically recognizes the small school issue and uses Russell's cutoff of at least 100 students for establishing growth targets for schools.¹⁹⁹ Schools with 11-99 students receive an asterisked API indicating the lower reliability of the index and too few students for establishing a valid growth target.²⁰⁰ Alternative accountability measures are being developed for small

¹⁹⁶ *Id.*

¹⁹⁷ Russell Report, *supra* note 108 at xii.

¹⁹⁸ *Id.* at xii.

¹⁹⁹ CEC § 52052(a)(2),(g).

²⁰⁰ See API reports, www.api.cde.ca.gov; Chart 11b: 2001-2002 API Growth Report.

schools and those that serve special populations.²⁰¹

Russell cites a series by the *Orange County Register* that questioned the accuracy of awards given to schools based on API growth.²⁰² In response, Rogosa argued that the probability was very high that each award school had made some gain.²⁰³ My read of this controversy is that the article authors viewed 5% as a modest growth target and wanted assurances *not* that the school had shown growth but that there was a high probability that the school had exceeded the required state growth target before being given an award.

Basically, Russell (and another group of researchers, Kane & Staiger) have raised three main issues related to error in the API accountability system:

- Potential advantage for small schools with larger sampling error;
- Potential disadvantage for larger schools with a greater number of numerically significant subgroups; and
- Potential unfairness of using two-year reporting cycles.

²⁰¹ CDE, *Alternative Schools Accountability Model Indicator Reporting Guide for School Year 2001-2002*, www.cde.ca.gov, July 2002.

²⁰² Campbell, R., *API's Error Margin Leaves a Lot to Chance*, *Orange Co. Register*, Aug. 11, 2002; Sharon, K. et al., *Test Scores Unreliable*, *Orange Co. Register*, Aug. 11, 2002; Sacchetti, M., *Awards Ignore Key Factors*, *Orange Co. Register*, Aug. 13, 2002.

²⁰³ Rogosa, D., *What's the Magnitude of False Positives in GPA Award Programs?*, www.api.cde.ca.gov, Sept. 9, 2002; Rogosa, D., *Application of OCR "margin of error" to API Award Programs*, www.api.cde.ca.gov, Sept. 9, 2002.

Rogosa demonstrates that all three concerns are unfounded.²⁰⁴ The latter issue was addressed above in demonstrating the difference between precision and sampling reliability and the lack of relationship between size of error and size of gains.

The data at the top of Table 15 address the issue of advantage for small schools (and consequent lack of incentives for large schools). These data indicate that when there is no actual improvement, smaller elementary schools are slightly more likely to win awards. But when true gains are large, smaller schools are substantially less likely to receive awards.

Similarly, regarding subgroup effects, Rogosa demonstrates that award probabilities are only slightly lower for schools with 4 numerically significant subgroups than schools with two.²⁰⁵ However, because larger schools are more likely to have more numerically significant subgroups, the subgroup effect may be offset by the size advantage.

More importantly, the use of the subgroup growth criteria in addition to overall growth targets reduces the number of false positives (schools that have not improved but receive awards).²⁰⁶ As the data in the bottom of Table 15 indicate, the expected number of schools that won awards but did not improve is only 97 out of about 3100 schools eligible for awards or roughly 3% of the total.

²⁰⁴ Rogosa, *supra* note 195 at Section 4.

²⁰⁵ *Id.*

²⁰⁶ *Id.*

Although a school could exceed an overall growth target by sampling error, it is much less likely that multiple subgroups will all exceed their targets by sampling error.

In addition, because the state rank ordered awards eligible schools by size of growth and gave the largest awards to the schools with the largest growth, it is unlikely that monetary awards were given to schools that had met their growth targets by chance good luck (unless, of course, there were adult testing irregularities, i.e., cheating occurred). Indeed, Rogosa demonstrates that contrary to the assertions in the *Orange County Register* article that 35% of schools receiving awards did not deserve them, only about 2% of the schools and 1% of the funds represented false positive results.²⁰⁷

In sum, the error arguments advanced by critics are based on incorrect statistical reasoning and confuse equality of opportunity with equality of results. As Rogosa has observed:

The accuracy of the measurement taken in isolation is not the basis for judgments about accountability systems, it's the use of the measure that matters.²⁰⁸

Reporting Errors

Russell asserts:

Since the current accountability system has been in place, other factors such as late delivery of tests to schools (50% of schools have reported this problem) and inaccurate

²⁰⁷ *Id.*

²⁰⁸ *Id.*

reporting of results for several schools have contributed to errors in measurement . . . ²⁰⁹

A report by the California State Auditor examined the issue of inaccurate reporting of results and concluded that confusion of responsibility between the *Board* and the *Department*, time pressures imposed by legislative deadlines, inadequate monitoring of the test contractor, absence of an implementation plan, and school district errors negatively affected the early years of the testing program.²¹⁰ Subsequently, several of its recommendations for improvement have been implemented.

California faced a significant challenge in implementing and administering a testing program with no statewide database and significant responsibility for data accuracy vested in the individual school districts. Considering the large numbers of students tested, it is not surprising that there would be "growing pains" in the initial startup of the program. However, district and school responsibilities for checking and verifying the accuracy of their data have been clarified, deadlines have been established for district signoffs, and consequently the number of errors and corrections has dropped dramatically. State staff has also become more knowledgeable about working with a testing contractor. Indications are that data problems are no longer the significant source of error they once were.

²⁰⁹ Russell Report, *supra* note 108 at xii.

²¹⁰ Noble, M., *STAR Program: Ongoing Conflict Between the SBOE & SPI as Well as Continued Errors Impede the Program's Success*, California State Auditor, April 2000.

Consequential Validity

Consequential validity refers to indirect effects, in addition to those of the test scores themselves, which are attributable to a testing program. Such indirect effects may be beneficial or detrimental. Standard 1.23 of the *1999 Test Standards* states:

When a test use or score interpretation is recommended on the grounds that the test or the testing program per se will result in some indirect benefit in addition to the utility of information from the test scores themselves, the rationale for anticipating the indirect benefit should be made explicit. Logical or theoretical arguments and empirical evidence for the indirect benefit should be provided. Due weight should be given to any contradictory findings in the scientific literature, including findings suggesting important indirect outcomes other than those predicted.

Comment: For example, certain educational testing programs have been advocated on the grounds that they would have a salutary influence on classroom instructional practices or would clarify students' understanding of the kind or level of achievement they were expected to attain. To the extent that such claims enter into the justification for a testing program, they become part of the validity argument for test use and so should be examined as part of the validation effort.²¹¹

Advisory groups to the *Superintendent* and the *Board* recognized at the outset the importance of consequential validity. In their initial recommendations for calculation of the first API scores in 1999, the *Advisory Committee* and *Tech Group* stated:

²¹¹ *Test Standards*, *supra* note 13 at 23.

A major priority of the accountability system must be to identify, evaluate, and mitigate unintended consequences.²¹²

Some researchers have purported to demonstrate negative consequences attributable to the introduction of high-stakes testing in states.²¹³ However, other researchers have criticized this work for over-interpretation of small differences, failure to demonstrate statistical significance, third-variable alternative explanations, and as premature due to limited implementation time.²¹⁴ Moreover, some researchers analyzing similar data have concluded that strong accountability programs are more helpful than harmful, particularly for minority students.²¹⁵

In addition to investigating potential negative consequences, it is also important when analyzing consequential validity to recognize positive consequences of testing and accountability programs. Specifically, in California, positive consequences of the accountability system, including the API, intervention program, and school rankings and awards, in-

²¹² PSAA Advisory Committee, *The 1999 Base Year Academic Performance Index (API): The Report of the Advisory Committee for the Public Schools Accountability Act of 1999*, Nov. 1999, p. 3.

²¹³ See e.g., Amrein, A. & Berliner, D., *An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing*, Arizona State University Education Policy Studies Laboratory, December 2002, www.edpolicylab.org.

²¹⁴ See Steinberg, L., *Does High-Stakes Testing Hurt Students? Read the Early Evidence with Caution* Education Week, February 5, 2003, p. 48.

²¹⁵ See study by Carnoy & Loeb cited in Viadero, D., *Researchers Debate Impact of Tests*, Education Week, Feb. 5, 2003, 1, 12. For example, these researchers found that states with stronger accountability programs made larger gains on grade 8 NAEP math from 1996 to 2000, especially at the proficient level. In addition, gains for African-American and Hispanic students exceeded those for white students. *Id.*

clude: (1) focus on teaching the content standards, (2) more efficient and targeted use of available resources, (3) improvement of low-per-forming subgroups, and (4) better decision-making.

Russell and other critics of the testing and accountability program allege negative consequences, including the teaching of test items and other unethical behaviors, failure of the tests to explain why schools performed poorly, increased dropouts and student retentions in grade, and improper classification of special education students. For example, Russell asserts:

[The CA accountability system] promotes practices that are of poor educational value. [These] questionable practices include ... investing time and resources in test preparation while decreasing or eliminating investments in non-tested [areas]; increasing retention without exposure to supplemental or alternative learning opportunities; aggravating school drop out rates; and increasing (often without sound reason) the number of students classified as having special educational needs.²¹⁶

However, these critics supply no reliable, credible evidence to support their claims of causation or expectations of explanatory power. Indeed, Russell admitted as much during his deposition.²¹⁷

Further, the state has policies and procedures for detecting and deterring the unethical behavior that concerns Russell. However, before ex-

²¹⁶ Russell Report, *supra* note 108 at vii.

²¹⁷ Russell Deposition, *supra* note 121 at p. 371, lines 15-22.

aming ways of attenuating Russell's alleged negative consequences, a word about the limitations of a major data source relied on by Russell followed by a discussion of the potential positive consequences of the California accountability system.

Limitations of Teacher Survey

When discussing consequential validity issues, Russell relies heavily on the results of a 2001 national survey of teachers conducted by the National Board on Educational Testing and Public Policy. But the survey has a number of serious limitations.

For the survey, teachers were selected by urbanicity, grade level, and subject area within each of 9 categories of states classified according to the stakes of their testing programs for schools and students. Thus, it was not a random sample of teachers in California and included only 433 California teachers who responded to that survey (about 0.1% of all California teachers).²¹⁸ No information is provided about the demographics of those teachers (e.g., ethnicity, years of experience) or the response rate of the survey. A low response rate, particularly without followup of nonrespondents, would cast serious doubt on the results.²¹⁹

Further, opinions from such a small sample of teachers with unknown

characteristics extracted from a larger survey sample with a different purpose probably do not provide a valid indication of what is happening in the approximately 8000 schools across the large state of California. Generally, those with strong opinions respond to such surveys and the amount of bias introduced by nonrespondents is unknown. The survey has not been published and I have been unable to obtain a copy.

Another source of school survey information is the ongoing evaluation of the California *High School Exit Exam* conducted by the Human Resource Research Organization (HumRRO or *external evaluator*). The *external evaluator* has studied a representative, longitudinal sample of 92 California high schools. Fifty-one percent of principals and 54% of selected teachers returned surveys in spring 2002.²²⁰ These data contrast with the Russell survey data as shown.²²¹

Topic	High School Evaluation		Teacher Survey (N=433)
	Principals (N=47)	Teachers (N=159)	
Alignment of district curricula	89%	85%*	62%
Alignment of textbooks	81%		40%

* covers ¾ to almost all; average of ELA & math

It appears that the California teachers selected for the national teacher survey were more pessimistic about alignment of curricula and textbooks than the high school principals and teachers surveyed by the *High*

²¹⁸ Russell Report, *supra* note 108 at 35; Walsh, M. & Sack, J., *Suits Contend Officials Fail to Obey ESEA*, XXII(21) EDUCATION WEEK 1, 13 (Feb. 5, 2003).

²¹⁹ Kerlinger, F. N. & Pedhazur, E.J. *Multiple Regression in Behavioral Research*, New York: Holt, Rinehart & Winston, 1973.

²²⁰ Wise, L. et al., *California High School Exit Examination (CAHSEE): Year 3 Evaluation Report*, Human Resources Research Organization (HumRRO), June 28, 2002, p. 74-75. Background information on the respondents and their schools can be found at 75-77.

²²¹ *Id.* at 81-82; Russell Report, *supra* note 108 at 36.

School Exit Exam external evaluator. This may be related to differences in survey methodology, differences in demographic characteristics of the participants or the fact that respondents in the evaluation study were offering opinions specifically on the *High School Exit Exam* with high stakes for students and schools while the Russell survey teachers were probably primarily considering the *Stanford Test* and California Standards Tests included in the API at the time of the survey.

The Russell survey respondents also appear to misunderstand the intent of the state assessments. Large numbers indicated that the state tests do not accurately measure what minority students and English language learners (ELLs) know and can do (85% and 96%, respectively).²²² While many of these students undoubtedly have useful and valuable skills not measured by the state tests, the purpose of the state tests is not to measure what students can do in general but rather to measure specifically whether students have learned the knowledge and skills in core academic areas specified in the California content standards. These other skills that ELL students may have cannot substitute for the tested skills. In most cases, ELLs' low test scores accurately indicate that they have not yet learned to demonstrate the tested skills in English.

In sum, the value of the teacher survey relied on by Russell to evaluate consequential validity issues is highly suspect. Keeping in mind

these limitations of the data on which Russell has relied heavily in asserting negative consequences of the state accountability system, the following sections consider in more detail the potential consequences of the API. Potential positive consequences are discussed first followed by alleged negative consequences.

Potential Positive Consequences

Teaching the Content Standards.
Russell acknowledges:

[W]hen high-stakes decisions are made based on test scores, teachers modify their instruction so that it focuses on the skills and knowledge included on the test, de-emphasizing skills and knowledge not on the test. It is reasonable to expect that teachers will "teach to the test" more often in schools that are performing poorly on tests used for accountability purposes given the close scrutiny such schools face.²²³

It is clear from the experience of other states that *what is tested is what gets taught*. Plaintiffs decry "teaching to the test" and they are correct if they are referring to inappropriate test preparation that teaches specific test items. But it is appropriate and desirable for teachers to refocus their efforts on teaching the knowledge and skills specified in the California content standards. If they do so, it will improve the achievement of all students on the academic content they are supposed to be learning and consequently improve their performance

²²² Russell Report, *supra* note 108 at 41.

²²³ *Id.* at x.

on the tested skills derived from those content standards.²²⁴

Historically, the purpose of public schools was to teach academic skills (remember the 3Rs – reading, writing, and ‘rithmetic?). In recent decades, this job has become diluted by cafeteria-style offerings and efforts to ameliorate the effects of social problems in the community. In response, the state has sought to refocus schools’ energies on ensuring that all students at least learn basic core subjects as delineated in the content standards. Other activities can be pursued as time permits but are no substitute for solid academic skills.

If poor and minority students in schools with social problems were excluded from the API, it would provide an incentive for these schools to ignore these students’ academic progress and to shift resources from teaching the state standards to other content or nonacademic activities. Especially for these students who have few out- of-school opportunities to learn core academic skills, failure to focus on the skills in the state content standards disadvantages these students in the present and in the future when they lack basic prerequisite skills to pass the *High School Exit Exam* or for more advanced work such as college prep classes.

Russell cites data from California respondents to the national teacher survey (with limitations described

²²⁴ See Steinberg, L., *Does High-Stakes Testing Hurt Students? Read the Early Evidence with Caution* Education Week, February 5, 2003, p. 48, 24 (“Does high-stakes testing encourage teaching to the test? Probably. But this is not a problem if the tests that teachers are teaching to are measuring things we want our students to learn.”).

above) indicating 80% report instruction in tested areas has increased and 58% report instruction in non-tested areas has decreased as a result of the state-mandated testing program.²²⁵ This probably is a good result for disadvantaged students who receive nearly all of their core academic learning in school and need those skills to compete effectively with their peers in the workplace, colleges and vocational training programs. Moreover, in that same limited study cited by Russell, about as many or more responding teachers indicated that instruction in physical education and foreign language had remained the same as said it had decreased.²²⁶

Russell also cites mission statements from several schools indicating schools have many and varied educational goals. He asserts that:

These are, arguably, all important aims for public education. However, they are outcomes ignored by California’s API-based accountability system.²²⁷

Nonetheless, it does not necessarily follow that all these goals are of equal importance or that it would be

²²⁵ Russell Report, *supra* note 108 at 37; see section on *Limitations of Teacher Survey*, *infra* at p. 58.

²²⁶ *Id.*

²²⁷ Russell Report, *supra* note 108 at 19-20. Coincidentally, Russell obtained his school mission statements through an internet search that required repeated random samples of 50 California schools to obtain the desired total of about 50 mission statements. Each time, internet mission statements were found for only 20%, 34% and 38% of the selected schools, respectively. *Id.* at 19. Apparently, no attempt was made to obtain written statements directly from the selected schools. With such a low response rate, there may have been a selection bias in the perspectives of the minority of schools with websites that included mission statements.

inappropriate for some electives and extra curricular activities to be judiciously replaced with targeted remediation when disadvantaged students have not yet achieved the state content standards in core subjects.²²⁸

In developing and adopting consensus content standards in core academic areas, California has created an educational priority for teachers to first teach and students to first learn what is contained in the academic content standards in the four core subjects before doing other things. If some teachers are not teaching (or do not want to teach) these consensus standards, one might wonder why what they are teaching (or want to teach) is more important than the California academic content standards and why their views should predominate.

Efficient and Targeted Use of Available Resources. The *Accountability Act* specifically provides that:

Schools are expected to meet [their API growth targets] through effective allocation of available resources.²²⁹

Lawmakers appeared to be indicating that they believed that schools could do a better job of allocating and managing their resources to provide students with instruction in core academic content. However, the legislature also recognized that schools face different challenges in this endeavor. To address this concern, the legislature created a similar

schools measure that focuses on inputs.²³⁰

The similar schools rankings provide each school with a measure of overall performance relative to the 100 schools with the most similar challenges on the input variables listed in the *Accountability Act*. For example, a school serving predominately disadvantaged students that has a state decile rank of 2 but a similar schools rank of 9, has students that are achieving better than most schools with similar inputs.

Nonetheless, if this school were not able to meet its API growth targets, it would be eligible to participate in the intervention program. A participating school receives additional resources to complete self-study activities, meet with an external evaluator, and draft an action plan to identify and change factors (including inputs) impeding its instructional effectiveness.²³¹

Improvement of Low-Performing Subgroups. In my experience with statewide testing, it is common for new statewide testing programs to draw criticism from advocates for "at risk" subgroups likely to score poorly on initial test administrations. However, over time, such programs typically have recorded substantial achievement gains for at risk populations that would not have occurred absent the public scrutiny and accountability associated with testing.

For example, initial passing rates for minority students on a high school

²²⁸ Nearly half of the respondents in Russell's limited teacher survey reported that "the state-mandated test has brought much needed attention to education issues in my district." Russell Report, *supra* note 108 at 41.

²²⁹ CEC § 52052(c).

²³⁰ CEC §§ 52056(a).

²³¹ See CEC §§ 52053-52056.

graduation test in Texas increased substantially over a four-year period. In 1994, 29% of African-Americans and 35% of Hispanics passed all tests taken for the first time in tenth grade. By 1998, these percentages had risen to 55% and 59%, respectively.²³²

Better Decision-Making. No test is perfect or completely error-free. But before deciding to eliminate a particular test use, one must consider the alternatives for decision-making. Without objective test information, decision-makers may be forced to rely on data that is less valid, less reliable, more prone to unidentified subjective biases, and less helpful for the intended purpose.

For example, some critics of testing have proposed the use of grades together with test scores in evaluating whether disadvantaged students have met a high school graduation test requirement. However, grades do not have the same meaning across classrooms because teachers weight factors such as attitude, effort, improvement, attendance and achievement differentially when assigning grades. For example, the U.S. Department of Education reports, "Students who earn mostly A's in disadvantaged schools achieved at the level of students earning mostly D's in affluent schools."²³³

Inclusion of such factors in student grades renders grades a poor substitute for tests designed to measure

²³² TEA Statewide Results, www.tea.state.tx.us.

²³³ U.S. Department of Education, Office of Educational Research and Improvement, cited by Achieve, Inc., *Staying on Course*, 2002, p. 3.

specific content area skills. While useful for certain purposes, dependence on teacher evaluations of students, in whole or in part, would produce data that is not comparable at the district or school level and may not even be comparable across sections of the same class. Attempts to aggregate such data would provide misleading, incomplete, and inaccurate measures of school effectiveness in teaching the California Content Standards.

Attenuating Alleged Negative Consequences

Detecting & Deterring Unethical Behavior. Russell asserts:

[I]ncreases in the [Stanford Test]-based API scores over the past few years may very well be the result of inferior, test-centered teaching practices as opposed to student improvement in terms of state standards.²³⁴

[T]he high stakes associated with some state-level testing programs leads to questionable educational practices such as focusing instruction on test-taking skills, . . . , altering test administration conditions, providing inappropriate instruction during testing, and, in some extreme cases, altering student response sheets.²³⁵

[T]here is clear evidence that schools are engaging in questionable practices to improve test scores. In the worst cases, these practices include outright cheating.²³⁶

With respect to the allegation regarding "test-centered teaching practices," deciding whether it is a problem depends on how that phrase is

²³⁴ Russell Report, *supra* note 108 at x.

²³⁵ *Id.* at 32-33.

²³⁶ *Id.* at 44.

interpreted. If it means teaching the domain of knowledge and skills sampled on the test, that is a good thing; if it means teaching the specific tested content and items, that is inappropriate test preparation.²³⁷ Inappropriate test preparation can be minimized with increased staff education efforts, increased test security, and investigation of suspicious circumstances.

The state tests are high-stakes for schools and that can lead some educators to seek unethical shortcuts to improved test scores. Detering such unethical behavior is an important state responsibility. California has procedures in place to monitor, investigate and sanction unethical testing practices.

Regulations governing API awards that were adopted by the *Board* on January 11, 2001 provide for invalidation of a school's API for a period of 1 or 2 years under the following circumstances:

- adult testing irregularities (e.g., changing answers on a student answer document) certified by the district;
- certification by the district that the API is *not* representative of the school's students;
- the APIs for the previous and current years are *not* comparable due to a substantial demographic change in the student population;

²³⁷ See also, Mehrens, W. & Kaminsky, J., *Methods for Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent?*, 8(1) EDUC. MEAS.: ISSUES & PRACTICE 14 (1989).

the proportion of parental testing waivers is 10% or more of the school's *Assessment Act* enrollment;²³⁸ or

credible evidence indicates that the integrity of the API was compromised (e.g., administering an alternate form of the *Stanford Test* prior to *Assessment Act* testing).²³⁹

According to *Department* data, 53 schools (about ½%) did not receive 2002 API growth results due to adult testing irregularities in 2001 or 2002. When a testing irregularity is suspected, the *Department* requests that the Superintendent of the District investigate the school involved.

The *Department* also uses erasure analyses to identify classrooms where further investigation is warranted. The criteria are:

- The total number of erasures in the class exceeded the state average by two standard deviations;
- The percent of students with erasures for the class exceeded 75%;

²³⁸ Note: For schools with parental waivers between 10% and 20%, statistical tests are performed by the *Department* to determine whether the schools' results are representative by grade level. The outcome of these tests determines whether or not the API is invalidated. CCR § 1032(d)(4).

²³⁹ CCR § 1032(d). See also Office of Legislative Counsel, Letter to Senator Alpert, Sacramento, CA, Sept. 28, 1998, www.cde.ca.gov/statetests/star/regs/. Emergency Awards Regulations provided for a 2-year penalty for adult testing irregularities or certification by the district that the API was not representative of a school. The *Board* later amended the regulations to permit schools to apply for a waiver reducing the penalty to one year when due diligence has been displayed by the district in reporting the incident to the *Department* AND only a small percentage of students in the school were involved in the irregularities. SBE, Minutes & Agenda Item #9, Nov. 8, 2000.

- For all tests taken by students in the class more than 75% of the erasures changed an incorrect answer to a correct answer; and
- For at least one subtest the percentage of erasures from incorrect to correct was greater than 90%.²⁴⁰

In 2002, the *Department* investigated 76 alleged irregularities. The type and frequency of irregularities and the *Department* actions taken are summarized below.

2002 Testing IRREGULARITIES	
	FREQUENCY
	18
	14
	10
	14
	2
	2
	1
	13
	2
	FREQUENCY
	14
	42
	20

The following is a sample cautionary statement used when posting results on the internet:

Grade 4 California English Language Arts Standards Test results should be interpreted cautiously due to a test examiner giving students incorrect directions for the writing portion of the test.

Russell provides only anecdotal support for his allegations of significant cheating. His limited teacher survey suggests that the three most

common test preparation activities reported by about 70% to 85% of the respondents are appropriate and desirable: "teaching test-taking skills, encouraging students to work hard and prepare, and teaching the standards [known] to be on the test."²⁴¹ Fifty-five to 65% use similar items or commercial test preparation materials, practices that may be acceptable as long as the skills in the content standards are being practiced, *not* the content of specific test items. Only 9% of the respondents reported using released test items when none have actually been released.²⁴² Exactly what this means is unclear. Perhaps the respondents meant practice or sample items.

Surprisingly, in Russell's limited teacher survey, 66% of the respondents indicated that "[s]tudents are under intense pressure to perform well on the state-mandated test."²⁴³ This is surprising because there are no state-mandated consequences for students on the *Stanford Test* or the *California Standards Tests*; only the *High School Exit Exam* has high-stakes for individual students and they are provided multiple opportunities to pass it. A little more than half of these same teachers report student lack of confidence when taking the state test.²⁴⁴ Perhaps the concerns of some educators and other close adults have adversely affected students' perceptions or teachers' perceptions of students' views.

²⁴⁰ CDE, Sample Letter to a District Superintendent, November 2002.

²⁴¹ Russell Report, *supra* note 108 at 38.

²⁴² *Id.*

²⁴³ *Id.* at 39.

²⁴⁴ *Id.*

Cheating on tests by adults or students is a serious violation and state accountability programs have a duty to minimize it so the program will be as fair as possible for all. However, just because some individuals may engage in unethical behavior is not a reason to abandon the testing or accountability program. It is a reason to be vigilant and to be proactive in sanctioning proven violators.

Allocating Responsibility for Explaining School Performance. Russell criticizes the API for not explaining why a low-achieving school has scored poorly. He asserts:

A system that focuses solely on student learning outcomes, no matter how broadly defined, cannot provide schools and their constituents with information that allows them to identify why students succeed or fail to succeed.²⁴⁵

Unless the State changes its system to provide information as to *why* schools perform as they do, it will never be able to target assistance in a rational way.²⁴⁶

In general, standardized achievement tests do not provide explanations for test performance for any entity, and there is no psychometric standard requiring that they do so. To provide detailed diagnostic information would require a prohibitively long test or complex data collection procedure using unacceptable amounts of instructional time. Given the practical constraints on test length and administration time for statewide tests designed to provide

²⁴⁵ *Id.* at vi, emphasis in original.

²⁴⁶ *Id.* at xi, emphasis in original.

individual scores for each student, these tests are only valid for making decisions about overall achievement of a domain of knowledge.

However, subarea scores (e.g., algebra, probability & statistics) and classroom item performance data (e.g., 80% of students in a class missed a math problem on finding the perimeter of a circular driveway) can be useful indicators of individual student and classroom strengths and weaknesses. It is the job of local educators to use this information to identify individual student and classroom weaknesses, to collect further diagnostic information to pinpoint specific learning deficiencies where weaknesses have been noted, and to implement appropriate and effective strategies for remediation.

This division of responsibility was specifically acknowledged in the *API Framework* adopted by the *Board*. It stated:

The use of local indicators, systems, and reporting for local uses should be encouraged in order to supplement statewide comparative and longitudinal information.²⁴⁷

In choosing effective remedial strategies, local educators must carefully evaluate all available information and potential alternative causes to determine why a particular skill was not learned. For example, a low-scoring English Language Learner may lack sufficient English language skills, substantive knowledge or both. The type and duration of remediation appropriate for this English Language Learner will de-

²⁴⁷ *API Framework*, *supra* note 32 at 4.

pend on correct identification of the cause(s) of the student's learning deficiencies.

Just as there are many factors that may negatively affect a student's achievement of academic skills, including, but not limited to, economic disadvantage, lack of parental involvement, poor attendance, low motivation, lack of appropriate instruction, substance abuse, family problems, illiteracy and limited English proficiency, so too are there many possible reasons for a school's lack of success. These factors do not invalidate the test results, but their identification can aid school staff in understanding why that school has not been effective in teaching their students the tested skills, and in formulating a locally-appropriate remedy.

Recognizing Lack of Causation in Dropout and Retention Rates. Russell uses enrollment data for 1998 9th graders and 2001 12th graders in the Los Angeles Unified School District to calculate a statistic he calls "imputed dropout rates" and concludes that they exceed 60% (the *Department* estimates about 25%).²⁴⁸ Using that data, he asserts:

[These d]ropout rates are important information because they could enable the state to ensure that improvements in test scores are not coming at the cost of having more students pushed out of school.²⁴⁹

However, he acknowledges the serious limitations of this procedure –

²⁴⁸ Russell Report, *supra* note 108 at 34.

²⁴⁹ *Id.*

lack of student level data and failure to account for student transience rates, migration rates, and retention rates. Nonetheless, he claims that his imputed dropout rates "provide a snapshot of what dropout rates might be."²⁵⁰ More likely, failure to consider such potent factors as mobility renders Russell's imputed dropout statistic meaningless.

Regardless of the actual dropout statistics, Russell presents no direct evidence that the *Assessment Act* tests or the API cause dropouts. He does, however, rely on indirect information from his limited teacher survey in which 23% and 33% of respondents, respectively, reported that state-mandated testing caused many students to drop out of high school or be retained in grade. Viewed the other way, these data also mean that significant majorities, 77% and 67%, respectively, did not believe the state-mandated test caused dropouts or retentions.

Since most *Assessment Act* tests have no stakes for individual students (the stakes are primarily for schools), it is equally possible that students who dropout do so for other reasons such as academic difficulties (e.g., failing required courses), the need to work to support the family, the need to care for siblings or sick parents, marriage/pregnancy, military service, vocational school, moving to another state, incarceration, etc. For example, a dropout study completed in Texas in the late 1990s indicated that nearly half of African-American and Hispanic dropouts were due to poor atten-

²⁵⁰ *Id.* at 34.

dance and/or low or failing grades.²⁵¹ As in Texas, students in California are ineligible for a high school diploma if they have not completed all their required course credits.

Similarly, if student retentions have increased as Russell alleges, it may be because teachers are becoming stricter in requiring skill mastery as a prerequisite to promotion to the next grade. That is, the tests may simply be confirming teachers' own judgments.

Even if some students dropout due to the challenge of higher standards, should a testing program designed to identify unsuccessful schools and students be abandoned because some students have given up? In Texas, data indicated that the number of minority students remediated as a result of the high school exit test far exceeded the number who may have dropped out due to the testing requirement.²⁵²

Dropping out is primarily a high school behavior. In California, the legal age for leaving school is 18. The only test in the California accountability system with high stakes for students is the *High School Exit Exam*. So far, the *external evaluator* has found no evidence of increasing

dropout rates in three consecutive years of evaluations.²⁵³

Following Legal Requirements for Special Education Classification. Recall that in his report, Russell claims that the California accountability system promotes "increasing (often without sound reason) the number of students classified as having special educational needs."²⁵⁴ However, Russell presents no evidence to support this claim and the available evidence contradicts it.

The Individuals with Disabilities Education Act (IDEA) mandates specific procedures that must be followed when identifying students for special education programs.²⁵⁵ Diagnostic evaluation by a licensed professional is required and parental permission must be obtained. Thus, a school cannot simply designate a student as special education when the state test is administered. Further, the IDEA requires special education students to be included in the regular state testing program to the maximum extent possible and with appropriate accommodations as needed.

The table below lists the percent of California general education students age 5-21 classified as special education from 1992 to 2001.²⁵⁶

²⁵¹ Texas Education Agency, *1996-97 Texas Public School Dropout Report*, Austin, TX, 1998 at 11.

²⁵² *Id.* at 5, 7, 11. Data indicated that in 1997, the ratio of: (1) the number of students known to have failed the graduation test in 10th grade but having been successfully remediated and passed the test by 12th grade, to (2) the estimated number of students who had the required course credits and *may* have dropped out due to the testing requirement, was approximately 54:1 for African-Americans and 56:1 for Hispanics.

²⁵³ HumRRO Year 3 Evaluation Report, *supra* note 220.

²⁵⁴ Russell Report, *supra* note 108 at vii.

²⁵⁵ Individuals with Disabilities Education Act (IDEA), 20 U.S.C. § 1400 et seq. (1997).

²⁵⁶ CDE, Special Education Programs in California: A Statistical Profile, Part I Student Population, www.cde.ca.gov, p. 1.

YEAR	PCT SPEC EDUC
	9.10
	9.35
	9.54
	9.62
	9.76
	9.82
	9.97
	10.11
	10.21
	10.12

The data in the table indicate that the percent of California students classified as special education has increased about 1% in the last ten years. However, the percent of students in special education was the same in 2001 as when the API was first introduced in 1999. Russell's claim of increased special education placements in California due to the introduction of the accountability system is refuted by these data.²⁵⁷

Consequential Validity Summary

An excerpt from a recent Newsweek article about consequential validity for ELLs summarizes the typical pattern of the early debates and later results surrounding the imposition of higher standards for historically disadvantaged groups.

Next, examine California's Proposition 227. Passed June 1998 by a 61 to 39 percent margin, it banned bilingual education in the state's schools. Educators widely opposed it; so did

²⁵⁷ Note that even if the number of special education students had increased significantly from 1999 to 2001, that would not by itself indicate that the accountability program was the cause. It would be equally plausible that newly recognized disabilities, improved identification procedures, greater parental awareness, mainstreaming policies, or an increase in the incidence of certain disabilities accounted for the change.

President Clinton. Prophecies of doom were widespread. Clinton said it would condemn immigrant children to "intellectual purgatory." The head of the San Francisco School Board said that "this would set our students back 30 years."

What happened? Test scores of children from Spanish-speaking families didn't drop. They rose. In second grade, average reading scores of students with limited English ability have jumped in the past two years from the 19th percentile nationally to the 28th percentile [35th in 2002]. In math, the same students went from the 27th to the 41st percentile [50th in 2002], according to The New York Times.

"I thought it would hurt kids," Ken Noonan, superintendent of schools in Oceanside, a city north of San Diego, told the Times. thirty years ago he helped found the California Association of Bilingual Educators. "The exact reverse occurred, totally unexpected by me," he said. "The kids began to learn – not pick up, but learn – formal English, oral and written, far more quickly than I ever thought they would."²⁵⁸

Additional qualitative information evaluating unintended consequences will be included in the final report of the external evaluator. The final report will consider both awards and interventions.²⁵⁹

Lessons From Other States

Russell asserts:

If asked to rank the quality and utility of the [accountability] systems in place in [MA, TN, TX, FL, MD, OH, MI, RI, KY, AL, & CA], the system currently in place in California [1999 Accountability Act] would be near the bottom of the list. The [API] it em-

²⁵⁸ Samuelson, R.J. *The Lesson of Tough Love*, Newsweek, September 2, 200, p. 27.

²⁵⁹ AIR Evaluation Report, *supra* note 78 at 87.

plays is simply incapable of providing the type of information the State policy-makers need to make rational decisions as to which schools need help and how to help them.²⁶⁰

Russell recommends Rhode Island's accountability system as a good model for California. He also recommends a model proposed but never implemented in Massachusetts.²⁶¹

Quality state academic content standards and assessments are the foundation of an effective accountability program. To evaluate the appropriateness and reasonableness of the Rhode Island model for California, a comparison of relevant outside evaluations and demographic data is useful.

Table 7 displays ratings of state standards and accountability programs by the Fordham Foundation and Education Week for selected states. The selected states are the 8 states receiving grades of B- or higher on their standards as judged by Fordham Foundation content experts. Of these states, California received the highest content standards grade awarded with an overall grade of A- and individual content area grades of A for *language arts*, history, math and science. For comparison, the U.S. average C- grade and Rhode Island's D+ are also included.

Among the states with standards ranked highest by Fordham, only Massachusetts received a slightly

higher combined grade for standards and accountability from Education Week (an A- to California's B+). Education Week also gave Rhode Island a D+.

Standards and assessments scores and ranks based on state evaluations by the American Federation of Teachers (AFT) in 2001 also indicate that California is doing well. The AFT awarded California the maximum number of points for clear, specific standards grounded in content resulting in a tied 1st place ranking for California among the states. California did not receive full credit on the alignment of their assessments to standards because Standards Tests for science and social science had not yet been implemented at the elementary and middle school levels. However, Rhode Island scored only 12 out of 24 points for standards and 6 out of 24 points for assessments resulting in rankings of 49th and 47th, respectively.

Fordham also classified states according to a combination of content standards grades and accountability. Table 8 presents these data using standards categories of solid (A or B average), mediocre (C average) and inferior or none (D or F average or incomplete), and accountability categories of strong and weak. With its top A- rating for content standards and rating of strong for its accountability program, California was classified among the 5 *Honor Roll* states described as "Only these 5 states can claim to be doing standards-based reform well." Rhode Island was classified in the lowest category

²⁶⁰ Russell Report, *supra* note 108 at v.

²⁶¹ *Id.* at 47.

of *Irresponsible States* described as “These 21 states cannot claim to embrace standards-based reform.”

Moreover, demographically California faces very different challenges than Rhode Island. Demographic data for the same selected states are summarized in Table 9. As indicated, California has more than 25 times the number of public schools and serves nearly 40 times as many public school students as Rhode Island. California’s students are 63% minority to Rhode Island’s 16% and 25% ELL to Rhode Island’s 7%. California’s annual expenditures are more than 25 times those of Rhode Island although Rhode Island has a larger percentage of disabled students and spends more per pupil than California. Of the 5 *Honor Roll* states, only Texas faces a demographic challenge similar to California.

Table 10 provides a comparison of test score data for these same selected states. Included are NAEP scores, state standards test scores, TIMSS scores, SAT scores, and ACT scores where available. Except in 4th grade reading, Rhode Island reports similar percentages of proficient or above students on the state test. Rhode Island has somewhat higher NAEP scores but the reading results date back to the beginning of the California accountability program and math scores are now 3 years old. Neither California nor Rhode Island participated in TIMSS. ACT composite and SAT verbal scores are similar for the two states but California’s SAT quantitative average is higher than Rhode Island’s for

high school students choosing to take those tests.

Russell further asserts that score gains are deceptive and cites Kentucky, where accountability assessment gains far exceeded NAEP gains.²⁶² However, Kentucky produced no individual student scores and its assessments were roundly criticized by two independent panels of national testing experts (one of which I served on). Kentucky also operated on the honor system and available evidence indicated that inappropriate testing practices were widespread.

In addition, Russell cites an electronic journal article by Haney as evidence of suspect gains on the Texas accountability test.²⁶³ However, this is a rehash of arguments presented in an expert witness report in the Texas graduation test litigation that were largely discredited by the judge who found that:

While the [graduation test] does adversely affect minority students in significant numbers, the [state] has demonstrated an educational necessity for the test and the Plaintiffs have failed to identify equally effective alternatives. . . . The [state] has provided adequate notice of the consequences of the exam and has ensured that the exam is strongly correlated to material actually taught in the classroom. In addition, the test is valid and in keeping with current educational norms. Finally, the test does not perpetuate prior educational discrimination Instead, the test seeks to identify inequities and to address them.²⁶⁴

²⁶² *Id.* at 24-25.

²⁶³ *Id.* at 25-26.

²⁶⁴ GI Forum, *supra* note 1.

Meanwhile, Texas has adopted a new, more rigorous testing program.

Public Opinion & External Evaluations

Russell presents Gallup polling statistics indicating that in 2001, 31% of the total public and 42% of minority respondents "believed there is too much emphasis on testing in schools."²⁶⁵ However, this means that the majority – nearly 70% of all respondents and nearly 60% of minority respondents – did *not* believe there is too much emphasis on testing in the schools.

Other evidence also indicates public support for standards and testing. The Third International Mathematics and Science Study (TIMSS) found:

What was "basic" in eighth-grade mathematics in the U.S. **differed greatly** in Japan and Germany. For example, in the U.S. the basic content included arithmetic, fractions and a relatively small amount of algebra. In Japan and Germany, the basic content included intense coverage of algebra and geometry – much more than in the U.S.

... This emphasis on the need to establish and implement international standards for U.S. curricula is supported not just by political leaders ..., educational leaders and professional groups, but also by business leaders. They too recognize the critical importance of having our schools produce students who can compete with their counterparts in an increasingly integrated and global economy.²⁶⁶

²⁶⁵ Russell Report, *supra* note 108 at 20.

²⁶⁶ U.S. National Research Center, Summary of Findings, TIMSS United States, Oct. 15, 1996, www.ustimss.msu.edu, emphasis in original.

Another group of researchers put it this way:

Challenging all students to meet common standards should be non-negotiable. These standards must be more than just minimum requirements; they must be anchored in the challenging content and skills that students need to succeed. The highest-performing school systems around the world use this formula of common standards and assessments. Students in these countries routinely outperform U.S. students on international assessments, not because they have more talent, but because their schools expect more from them.

If these international comparisons are not convincing enough, we can find plenty of other evidence of the need for common, high standards. Too many students graduate from high school unprepared for the challenges that lie ahead. Increasing numbers of students at four-year colleges need remedial education in reading, writing, or mathematics. Employers tell a similar story: 34% of job applicants tested by major U.S. firms in 2001 lacked sufficient reading and math skills to do the jobs that they sought.²⁶⁷

California Opinion Data

The results of a random-sample, telephone survey of 800 frequent California voters conducted in August 2002 by Fairbank, Maslin, Maulin & Associates for the California Business for Education Excellence Foundation are presented in Table 11. The margin of error was $\pm 3.5\%$.

Three quarters of respondents believed there are ways to hold schools

²⁶⁷ Gandal, M. & McGiffert, L. *The Power of Testing*, 60(5) EDUCATIONAL LEADERSHIP 39 (Feb. 2003), citations omitted.

accountable for student progress, 78% favored the use of standardized tests, and 80% said they were more included to favor standardized testing because public reporting of standardized testing results puts pressure on schools to do better.

The sample was split on the question of whether it is more important to know if students are learning state standards or how California students are doing compared to students nationally. Having the same learning standards for all schools statewide was endorsed by 86% of respondents and 58% agreed that the momentum should be kept going with more reform.

Achieve summarizes public support for standards-based reform:

There should be no doubt that the public firmly supports using common, challenging standards to raise student achievement, measuring results, and holding schools and students accountable for performance. Polls taken over the last five years consistently make clear that the public agrees that states have the proper strategy.²⁶⁸

Teacher Variables

Teacher quality matters. As Russell acknowledges:

When students are repeatedly exposed to low-quality teaching, their learning suffers.²⁶⁹

²⁶⁸ Achieve, Inc., *Staying on Course*, 2002, p. 17.

²⁶⁹ Russell Report, *supra* note 108 at ix.

The difficulty (and sometimes disagreement) lies in specifying what constitutes quality teaching.

The *No Child Left Behind* federal law requires highly qualified teachers in all classrooms serving Title I students by the 2005-06 school year.²⁷⁰ States propose and the administration reviews their definitions of highly qualified teacher.

California originally proposed that "teachers with nonclassroom work experience be counted as *highly qualified* ... provided that their previous work could be construed as relevant to their teaching fields, and that they were making progress on other certification requirements" and it was rejected by the U.S. Department of Education.²⁷¹ The state is currently working on a revised definition that will be submitted for review this spring. Legislation that would strengthen teaching requirements is also under consideration.

Correlation with SES

Russell asserts:

[T]here is a clear relationship between the percentage of emergency credentialed teachers within a school and API scores – as the percentage of Emergency Credentialed Teachers increases, API scores decrease. ... While several factors combine to influence the relationship between SES and API scores, teacher quality (as

²⁷⁰ NCLB, *supra* note 186 at § 1119 ("[E]ach State educational agency receiving assistance under this part shall develop a plan to ensure that all teachers teaching in core academic subjects within the State are highly qualified not later than the end of the 2005-2006 school year.")

²⁷¹ Walsh, M. & Sack, J., *supra* note 218 at 13.

represented by Emergency Credentialing) is one key factor.²⁷²

It only makes sense, then, that for schools that have a high percentage of emergency credentialed teachers, interim goals should focus on decreasing the percentage of emergency credentialed teachers (ideally to 0%) rather than on increasing students' test scores. Only after significant progress towards this interim goal has been reached should attention turn to changes in test scores.²⁷³

The Russell report takes the simplistic view that because school API scores and percents of emergency credentialed teachers are correlated, the state can improve academic performance by requiring schools to reduce the number of non-fully credentialed teachers over an unspecified period of time. Meanwhile, schools would not be held accountable for student outcomes. He presents the following correlations to support his argument.²⁷⁴

	% Teach Emerg Cred	% Free/Reduced Lunch	% Parents NOT HS Grads
API	-.46	-.81	-.73
% Teach Emerg Cred	—	.36	.34
% Free/Red Lunch		—	.75

However, these data indicate (and he acknowledges) that API scores are also correlated with SES variables (percent free/reduced lunch and percent of parents not high school graduates). In fact, the relationship between the SES variables and API scores is much stronger than the relationship between API

²⁷² Russell Report, *supra* note 108 at 46.

²⁷³ *Id.* at xx.

²⁷⁴ *Id.* at 46.

scores and emergency credentials. The SES variables account for 53%-66% of the variance in API scores while the percent of teacher emergency credentials accounts for only 21% of the variance in API scores.

Thus, although none of these correlations can establish a cause and effect relationship, these data suggest that the percent of teachers with emergency credentials is only a minor factor in API scores while SES is a major factor. Most importantly, these relationships do not prove that low API scores are caused by too many teachers with emergency credentials. Indeed, in a memo to Russell, his research assistant stated:

[I]t appears that the effect of emergency credentialing is ... much smaller than the simple correlations would lead you to believe once other basic school/student characteristics [e.g. SES] are [modeled].²⁷⁵

Rogosa has addressed the "correlation is NOT causation" problem of reporting percents of emergency credentialed teachers for high and low-scoring schools and concluding that reducing the former would fix the latter. These typically reported data are presented on the left side of Table 12 by API decile. Clearly, schools in the lower deciles have a higher average percent of emergency credentialed teachers and lower average percent of fully credentialed teachers than schools in the upper deciles. The relationship is monotonic and strong.

²⁷⁵ Memo from Raczek to Russell, Jan. 11, 2002, PLTF-XP-MR 1096.

However, as Rogosa observed:

[I]t's obvious that the schools that draw from advantaged student populations tend to be those schools whose pay, resources, and working conditions are attractive to teachers (and thus have few emergency credentialed teachers). Conversely, schools that draw from disadvantaged student populations are also those schools whose pay, resources, and working conditions are far less attractive to teachers, and these schools by necessity have far more emergency credentialed teachers.

...
Table [12] in no way implies that if low scoring schools were instantaneously transformed to have no emergency credential[ed] teachers that students would be better off (at least in terms of test performance).²⁷⁶

A more accurate picture of the relationship between emergency credentialed teachers and student performance can be obtained by examining performance for students with more similar initial levels of performance. The center portion of Table 12 reports the gain in 2001 API scores by decile for SES_D students in schools reporting no emergency credentialed teachers compared to schools with more than 15% emergency credentialed teachers (the state average is 11%). On the right side of Table 12, the same comparisons are reported for SES_D students in schools with high (more than 50%) SES_D enrollments.

For both of the SES_D comparisons, there is no clear pattern of advantage. Across the decile range, about half the time API gains are greater in

²⁷⁶ Rogosa, D., *Teacher Credentials and Student Progress: What do the data say?*, Stanford University, December 2002, p.1-2.

schools with higher percentages of emergency credentialed teachers. Conversely, if the percent of emergency credentialed teachers had a significant impact on API growth, one would expect the results for the schools with no emergency credentialed teachers to be consistently better and they are not. Rogosa concluded:

[T]he main point is that from these [2001] data there's little indication that reducing [the number of emergency credentialed teachers] would be notably beneficial for student performance.

...
[Data for 2000 did] show a somewhat consistent 8-10 point [API] advantage for [schools with no emergency credentialed teachers]. But 8 to 10 points is not a large effect: a little less than each student getting one more question correct on the [*Stanford Tests*]. Or to put it in a NCLB metric, 8 to 10 API points approximately represents one additional percent of the students above proficient, a fraction of the mandated annual yearly improvement.

...
It is what these [data] don't show that is the most important point – these [data] do not reveal a large systematic advantage for students in schools with [no emergency credentialed teachers].²⁷⁷

Data for Schools Attended by Named Plaintiffs

Of the allegations made by Plaintiffs regarding deficiencies in teachers, instructional materials and facilities in specific schools attended by named plaintiffs, teacher inadequacies may be the most potent factor in API performance. Though inconven-

²⁷⁷ *Id.* at 3-4.

ient, students can share books, use copied materials or internet resources, wear coats in a cold classroom, or use a restroom on another floor. But if the classroom teacher is not able to effectively focus instruction on the state content standards for the subject area of the class, disadvantaged students may be ill-equipped to learn the material on their own.

Thus, of all the schools named in the Plaintiffs statement of liability, the ones attended by named plaintiffs where it is alleged that "class representatives have suffered due to lack of access to qualified teachers"²⁷⁸ might be most likely to be unsuccessful in meeting their annual API growth targets. They might also be expected to have the greatest number of teachers with emergency credentials and the highest student/teacher ratios.

For the schools listed in the teacher quality section of the Plaintiffs' Liability Statement referenced above, and based on data available on the *Department* API website, Tables 13a-c and 14 summarize API results for the 1999-2000, 2000-2001 and 2001-2002 API reporting cycles and 2002 demographic characteristics for 6 elementary schools, 5 middle schools, and 5 high schools listed as attendance sites for the named plaintiffs alleging poor quality teachers. Charts 19 and 20 summarize awards and intervention funding and percents of full and emergency teacher credentials for these same schools.

²⁷⁸ *Williams v. State of California*, Case No. 312236, San Francisco Superior Court, Plaintiffs' Liability Disclosure Statement, Oct. 3, 2002, p. 56.

Several trends are apparent in these data.

1 API Growth, Awards & Intervention. First, 13 of these 16 schools rank in the first or second decile statewide but range from 1-7 in their similar schools decile rank. That is, some of these schools are doing much better than others relative to schools with similar demographics and input challenges. The elementary schools have met nearly all of their growth targets, middle schools have been successful about half the time and the high schools have rarely met their growth targets.

All but one of the 6 elementary schools have received monetary awards totaling \$620,052. Two of the 5 middle schools have received a total of \$172,022 and one of the 5 high schools received \$386,127 in monetary awards.

Two of the elementary schools received a total of \$390,600 in state intervention program funding in 2003, one as a new grant and the other as a continuation grant. Four of the five middle schools received new intervention program funding and the other continuation funding totaling \$1,967,000. Three of the five high schools received new intervention program funding in 2003 and one received continuation funding totaling \$2,709,600.

Of a total of 16 schools, the 5 schools not receiving intervention funding have all received monetary awards. Overall, these 16 schools have received \$1,178,201 in monetary awards in the last three years.

They have also received \$701,600 in continuation funding and \$4,365,600 in new funding from the intervention program in 2003. The 3 schools with continuation funding received grants the previous year and the 8 schools with new intervention grants in 2003 should receive funding again in 2004. In sum, these 16 schools have received a total of \$6,245,401 in additional state monies they would not have received if there had been no API accountability program.

2 Emergency Credentialed Teachers. Second, the percent of teachers with emergency credentials varies widely among these 16 schools, ranging from 2% to 64%. Of the 8 schools with percents of emergency credentialed teachers substantially exceeding the state average of 11%, 5 received monetary awards and one became eligible in 2002. Of the 8 schools at or below the state average, 3 received monetary awards and one became eligible in 2002.

Of the 11 schools receiving intervention program funding, 5 substantially exceeded the state average percent of emergency credentialed teachers and 6 did not. Of the 5 schools with similar schools ranks above the 5th decile, 60% substantially exceeded the state average for emergency credentialed teachers. In sum, for this group of 16 schools alleging teacher quality problems, the percent of emergency credentialed teachers is not a good indicator of relative success.

3 Average Class Sizes. Third, the average class size varied across schools but was also not re-

lated to success. Of the schools with the largest and smallest class sizes within each level (elementary with a tie for largest, middle, and high school), $\frac{3}{4}$ of the schools with the largest class sizes received awards but only $\frac{1}{3}$ of the schools with the smallest class sizes received awards. One of the three schools with the smallest class sizes had a similar schools rank above the 5th decile while 2 of the 4 schools with the largest class sizes did. Almost all of these schools have large proportions of SES_D or ELL students or both.

CONCLUSION

The *Assessment and Accountability Acts* indicate that California is seeking the same change in results as the federal *No Child Left Behind* mandate – nearly all students proficient in core academic subjects. As Hanushek indicated in the Kansas City case, this requires incentives that make student achievement matter to schools. This is not likely to happen if the state is treated like a giant research lab where the main goal is to collect and study input data.

When a school has not met its growth target for academic achievement by its students, there are many factors that may have contributed to that result. The explanations for one school may be quite different than the explanations for another. Just as some students are successful in overcoming adverse conditions in low-performing schools and scoring well, satisfaction of a set of arbitrary

criteria for per student square footage or numbers of available instructional materials in high performing schools does not guarantee student success.

The *Intervention Program* detailed in the *Accountability Act* legislation prescribes a two-step process for dealing with unsuccessful schools. This two-step approach assigns primary responsibility to districts and schools for meeting performance goals and secondary responsibility to the state in the event of repeated lack of progress over time. This approach avoids having a "one-size-fits-all" approach with additional arbitrary regulations that could unduly interfere with the operation of successful schools or prevent an unsuccessful school from working creatively with the local community to solve its problems. Rather than using limited state resources to police everyone, available state resources are concentrated incrementally where they are most needed.

The API is a work in progress. California has been careful to include new measures only when valid and reliable as required by statute. Within a few more years, the API will include *Standards Tests* in all core subjects completely aligned with state academic content standards. Work is progressing on a student database that could support the inclusion of nonacademic indicators in the API.

It would be unfortunate to abandon the progress that has been made in student achievement of state content standards because some schools

have not met their responsibilities in other areas. It would also divert attention from the primary mission of public schools to teach the state academic content standards if measurement of academic achievement was put on hold while some low-performing schools worked to correct input deficiencies.

California wants to encourage schools to use their limited resources to provide all their students with the best possible instruction on the state academic content standards. Annual measurement of student achievement and reporting of results using an accountability measure such as the API is a reasonable way for the state to achieve this goal.

The API accountability system did not create the social problems faced by ethnic and *SES_D* subgroups but is contributing to their improvement. It has provided much needed incentives to improve the achievement of disadvantaged students with some schools already demonstrating significant improvements.

All of the schools attended by named plaintiffs alleging teacher problems have received significant additional state funds as awards or intervention funding that would not have been available without the API and the state academic accountability program. The 11 of 16 schools receiving intervention funding are required to study input measures, determine with the help of an external evaluator why the school has been unsuccessful, and decide how best to use the additional funds to address input deficiencies.

It will take time to build a system in which graduating seniors will have come through 12 years of standards-based instruction and had the prerequisites to tackle challenging high school work. It would be unfortunate if the state were required to revert back to the unsuccessful Kansas City model before the API accountability system has had a full opportunity to be effective.



S.E. Phillips
Consultant
Box 384
W. Paducah, KY 42086

CHART 1
Overview of the California Academic Accountability System

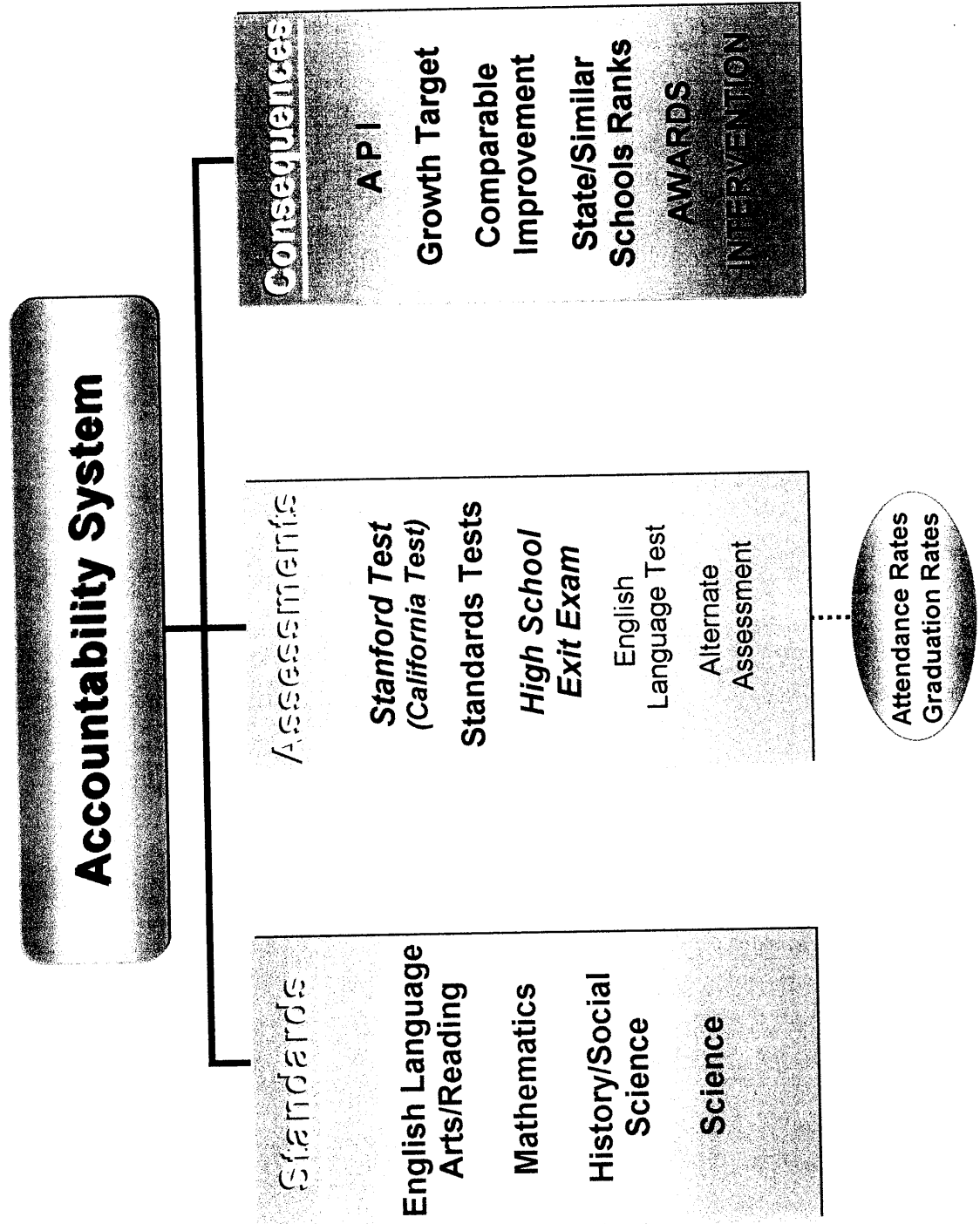


CHART 2
Organizational Chart for the California Academic Accountability System

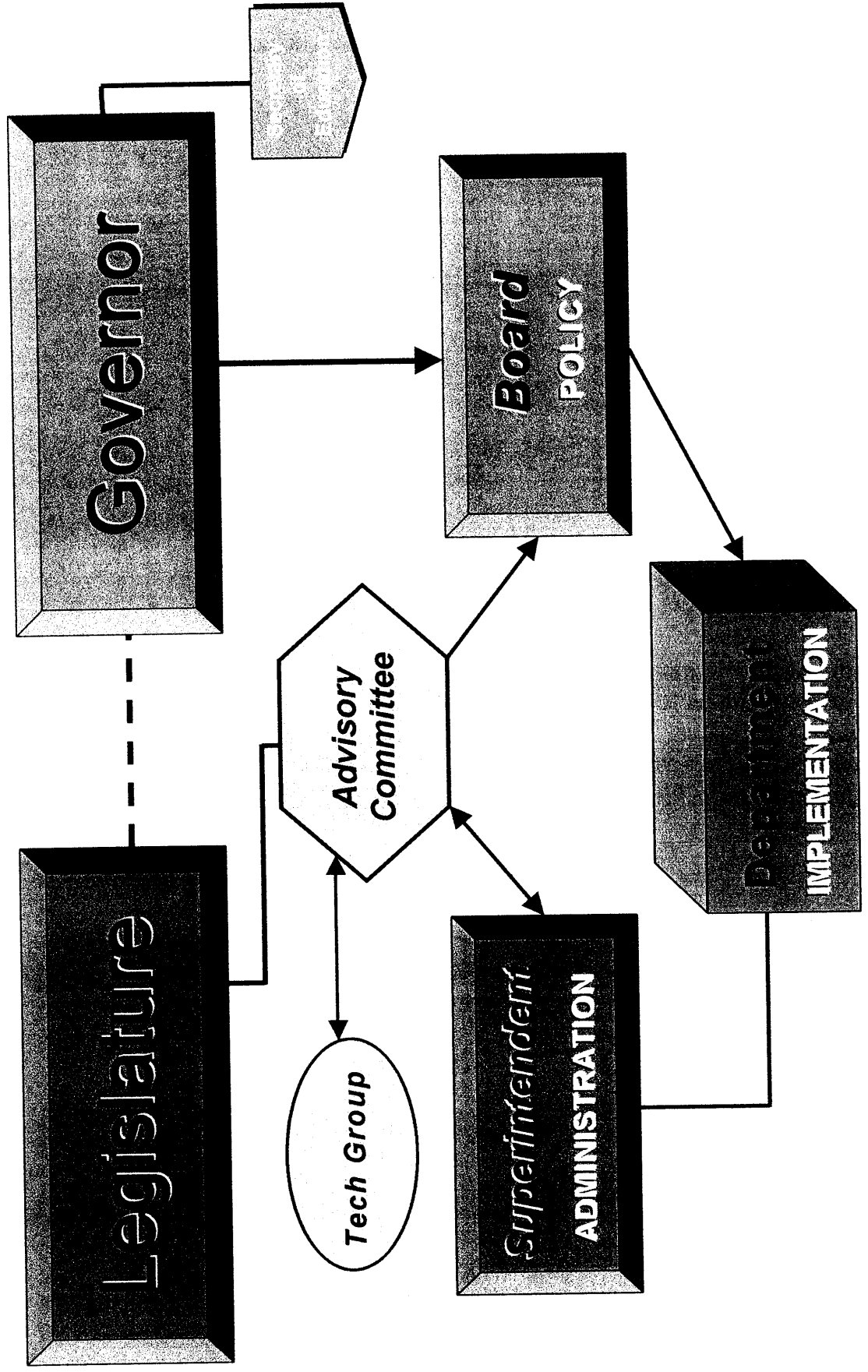


CHART 3
Timeline of Board Actions Involving the API

1999	<u>Item# & Type</u>	<u>Content</u>
Jul 15	26. Action	Approved the API Framework
Oct 7	20. Information	Report from <i>Advisory Committee</i> re 1999 Base API
Nov 9	23. Action	Approved 1999 Base API Calculation Recommendations of <i>Advisory Committee</i> (except exclusion of ELLs)

2000		
May 11	12. Information	<i>Department</i> Presentation re Coordination of Statewide Tests
Jun 7	9. Information	<i>Tech Group</i> Presentation on the API
Jul 12	11. Action	Approved <i>Stanford Test</i> as only Component of 2000 API; Approved Awards Program Recommendations of <i>Advisory Committee</i> (with 95% Elem/Mid & 90% High School Participation Rates) Approved Alternative Accountability System Recommendations of <i>Advisory Committee</i>
Oct 11	12. Action	Adopted Draft Emergency Regulations for API Award Programs
Nov 8	9. Action 12. Action	Approved Revised API Awards Regulations re School Waivers for Testing Irregularities & Parent Exemptions & Initiated Proposed Rulemaking

2001		
Jan 11	35. Information 36. Action	Held Public Hearing on API Awards Regulations Approved Amendments to Emergency Regulations for API Awards Program re Data Collection Deadline & Parent Opt-out Provisions
Feb 7	16. Action 17. Action	Approved 2001 Growth API Policy Recommendations of <i>Advisory Committee</i> with Reduction of Parent Exemptions to 10% & 15-day Review of Revised API Award Program Regulations

2001 Cont'd

Mar 7	21. Action	Adopted Permanent API Award Program Regulations
May 9	10. Information	<i>Department</i> Presentation on Proposed Procedure for 2 nd Year Awards Eligibility for Schools with Excessive Parent Exemptions in 2000
Jun 6	8. Information 9. Information	<i>Tech Group/Advisory Committee</i> Presentation on Adding the ELA Standards Test to the API & Discussion of <i>Department</i> Draft Proposal for Calculation of District APIs
Jul 11	11. Action	Amended API Award Program Regulations re Reduction of Parent Exemptions to 10% & 85% Minimum Participation per Subtest
Sep 5	17. Action	Approved Addition of ELA Standards Test & Revised Component Weights for 2001 Base API
	Information	Discussed Proposed District API Reports
	18. Action	Approved Amended API Award Program Regs
Oct 10	11. Information	Held Hearing on Proposed Amendments to API Award Program Regulations <i>Department</i> Presentation on 2001 API Awards
Nov 7	17. Action	Adopted Amended API Award Program Regs

2002

Mar 6	27. Action	Approved Addition of Mathematics Standards Test & Revised Component Weights for 2002 Base API
Apr 24	9. Information	Reviewed Six-Year Plan for API Development
May 30	31. Information	Discussed Integrating the <i>High School Exit Exam</i> into the 2002 Base API
	32. Action	Adopted the High School History Standards Tests as Sole Social Science Component in 2002 Base API
Jun 26	5. Action	Approved Addition of <i>High School Exit Test</i> & Revised Component Weights for 2002 Base API
Oct 9	4. Information	Discussed the NCLB Accountability Plan

2003

Jan 8	3. Action 5. Action	Approved an NCLB Accountability Plan for AYP Approved Revisions in the 2002 Base API Weights
-------	------------------------	---

CHART 4
API Policy Reports Prepared for the Board
by the Advisory Committee & Tech Group

<u>Date</u>	<u>Title</u>
Jun 1999	Framework for The Academic Performance Index
Nov 1999	The 1999 Base Year Academic Performance Index (API)
Jul 2000	The Alternative Accountability System
Nov 2000	Recommendations on Waivers of the Awards Regulations: Adult Testing Irregularities and Parent Requests for Exemptions
Feb 2001	Summary of Policy Issues Related to the 2001 Growth API Issue Paper: Parental Waivers and the 2001 Growth API Issue Paper: Percent Tested in a Content Area and 2001 Growth API Issue Paper: Deadlines for Data Submissions and Corrections for the 2001 Growth API and the Awards Programs Issue Paper: Testing Irregularities and Academic Performance Index (API) Validity Issue Paper: Participation Rates and Academic Performance Index (API) Validity
Jul 2001	The 2001 Base Academic Performance Index (API): Integrating the California Standards Test for English-Language Arts into the API
Aug 2001	The District Academic Performance Index
Mar 2002	The 2002 Base Academic Performance Index (API): Integrating the California Standards Test in Mathematics into the API
Apr 2002	The Academic Performance Index (API): A Six-Year Plan for Development (2001-2006)*
May 2002	The 2002 Base Academic Performance Index (API): Changes in the High School Social Science Indicator
May 2002	The 2002 Base Academic Performance Index (API): Integrating the California High School Exit Examination (CAHSEE) Results into the API
Jan 2003	Revisions in the 2002 Base Academic Performance Index (API)

* Prepared by the Department with input from the Tech Group and Advisory Committee.

CHART 5
Sample 2000 API Elementary School Calculation*

Stanford Test		Reading		Language		Spelling		Mathematics	
A	B	C	D	E	F	G	H	I	J
Performance Bands	Weighting Factors	Percent of Pupils in Each Band	Weighted Score in Each Band (B x C)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x E)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x G)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x I)
80-99 th NPR	1000	5%	50	10%	100	5%	50	5%	50
60-79 th NPR	875	5%	44	10%	88	10%	88	10%	88
40-59 th NPR	700	25%	175	30%	210	25%	175	25%	175
20-39 th NPR	500	35%	175	30%	150	35%	175	35%	175
1-19 th NPR	200	30%	60	20%	40	25%	50	25%	50

Total Weighted Score Across Bands	504	588	538	538
Content Area Weight	x .30	x .15	x .15	x .40
Total Weighted Score for Content Area		88	83	
		+	+	+
				=
				535

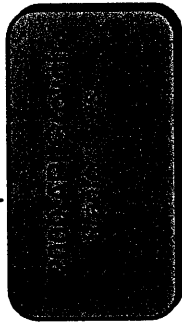
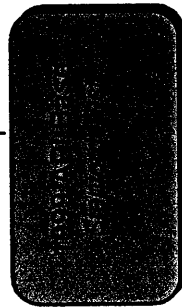
2000 API

* Adapted from: CDE, 2000 API Base Report Information Guide, January 2001 at 18.

CHART 6
API Reporting Cycles*



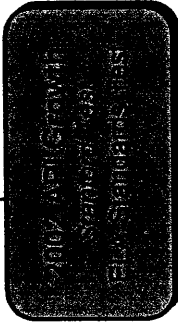
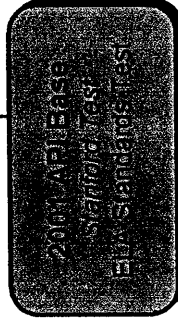
┌ 1999 to 2000 Growth ┐



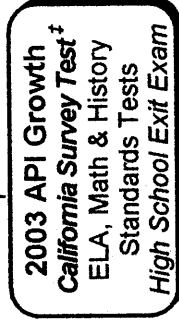
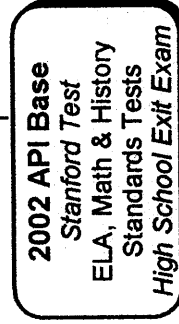
┌ 2000 to 2001 Growth ┐



┌ 2001 to 2002 Growth ┐



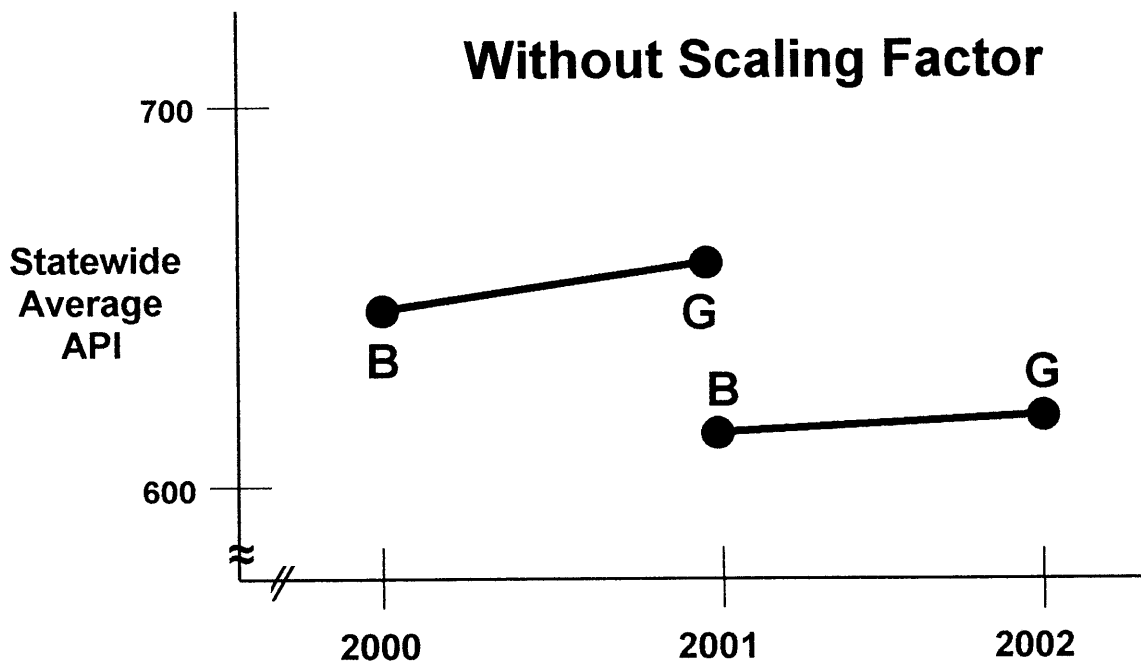
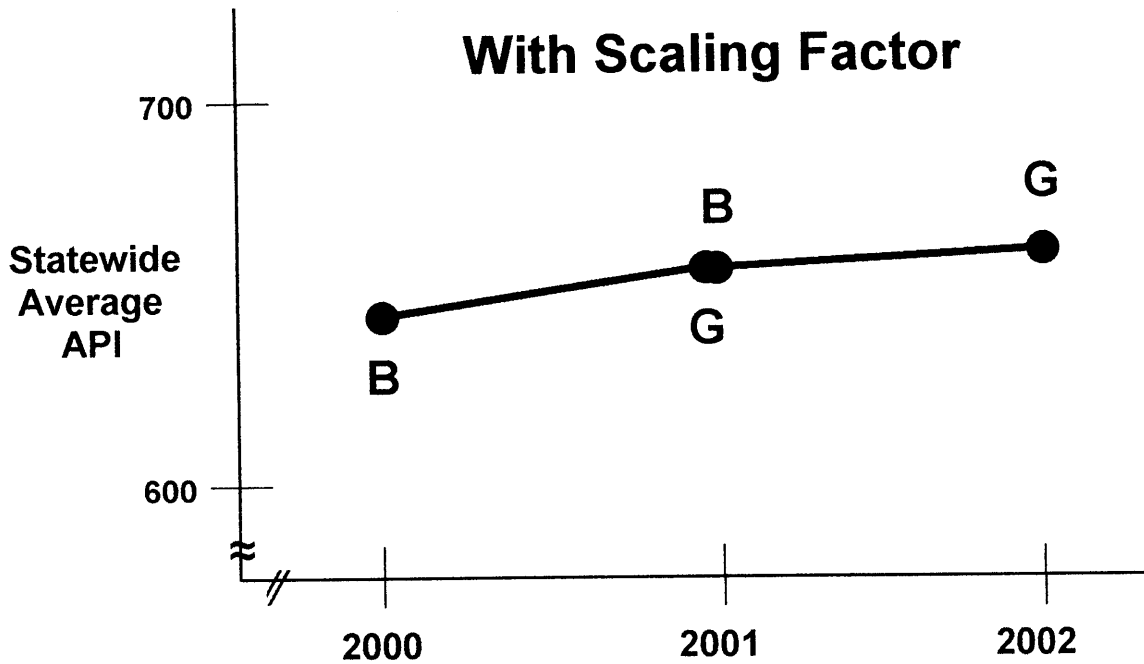
┌ 2002 to 2003 Growth ┐



* Adapted from CDE, 2000 API Base Report, January 2001 at 14.

† Linked to the Stanford Test.

CHART 7
Purpose of the Scaling Factor in the API*



* B = Base; G = Growth; Green = 2000-2001; Blue = 2001-2002; Adapted from: SBE Minutes & Agenda Item #17, September 5, 2001, Attachment 9, p. 12.

CHART 8

Sample 2002 Base API Calculation for a High School*

Standards Test		Language Arts		Mathematics		History		ELA Exit Exam		Math Exit Exam	
A	B	C	D	E	F	G	H	I	J	K	L
Performance Levels	Weighting Factor	Percent of Pupils in Each Level	Weighted Score in Each Level (B x C)	Percent of Pupils in Each Level	Weighted Score in Each Level (B x E)	Percent of Pupils in Each Level	Weighted Score in Each Level (B x G)	Percent of Pupils Passing	Weighted Score in Each Level (B x J)	Percent of Pupils Passing	Weighted Score in Each Level (B x L)
Advanced	1000	8%	80.00	9%	90.00	11%	110.00	75%	750.00	64%	640.00
Proficient	875	23%	201.25	20%	175.00	24%	210.00				
Basic	700	35%	245.00	32%	224.00	28%	196.00				
Below Basic	500	21%	105.00	23%	115.00	19%	95.00				
Far Below Basic	200	13%	26.00	10%	20.00	18%	36.00				
Untested	200	0%	0.00	6%	12.00	0%	0.00				
		667.25 x .36		636.00 x .18		647.00 x .20		800.00 x .10		712.00 x .05	
		+		+		+		+		+	

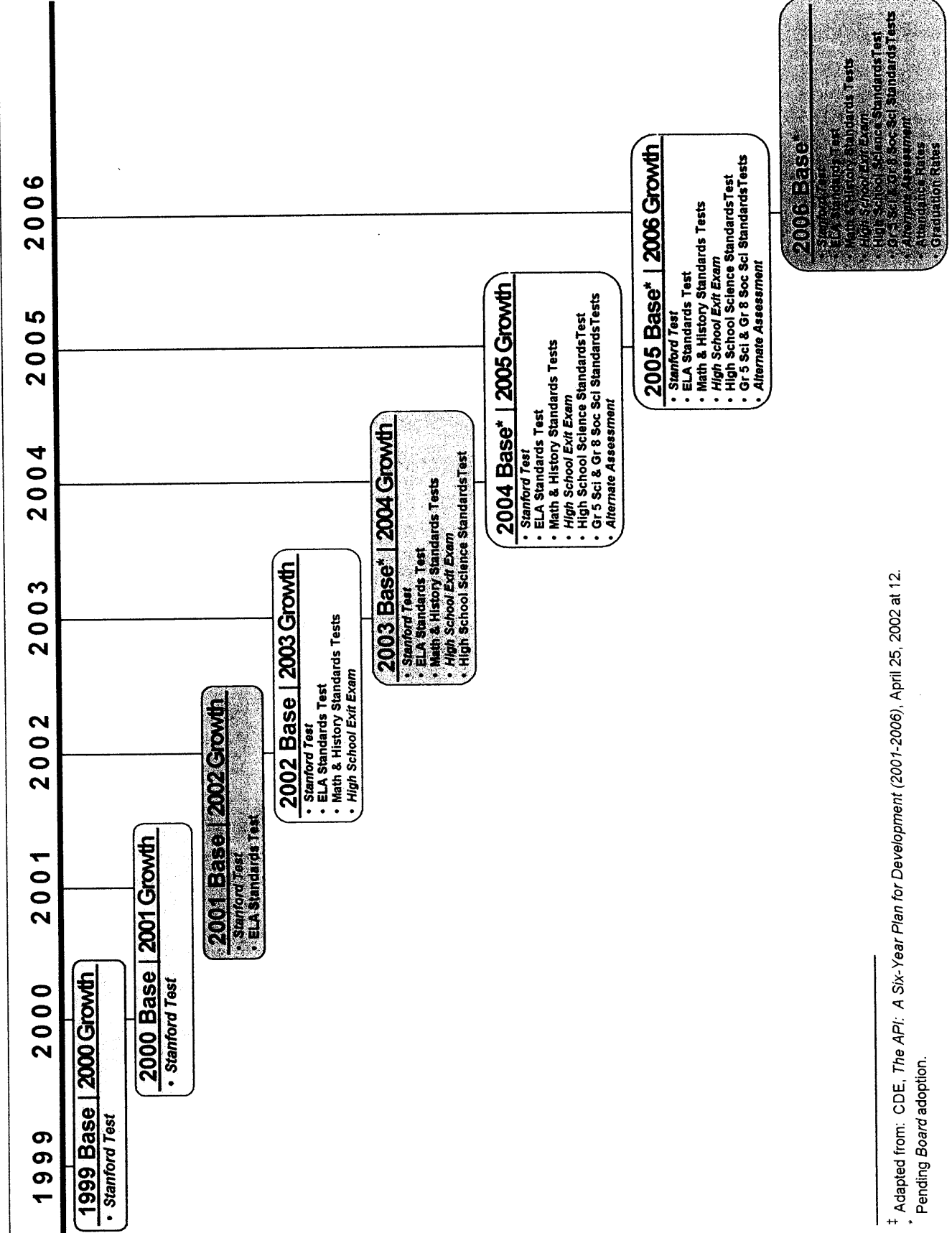
Stanford Test		Language Arts		Mathematics		Science			
A	B	C	D	E	F	G	H	I	J
Performance Bands	Weighting Factor	Percent of Pupils in Each Band	Weighted Score in Each Band (B x C)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x E)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x G)	Percent of Pupils in Each Band	Weighted Score in Each Band (B x J)
90-99 th NPR	1000	9%	90.00	12%	120.00	21%	210.00	14%	140.00
60-79 th NPR	875	17%	148.75	26%	227.50	21%	183.75	22%	192.50
40-59 th NPR	700	23%	161.00	23%	161.00	20%	140.00	22%	154.00
20-39 th NPR	500	23%	115.00	22%	110.00	19%	95.00	21%	105.00
1-19 th NPR	200	28%	56.00	17%	34.00	19%	38.00	21%	42.00
		570.75 x .03		652.50 x .03		666.75 x .03		633.50 x .03	
		+		+		+		+	

2002 Base API → **669**

Scale Calibration Factor = **669**

* Adapted from: Advisory Committee, *The 2002 API: Integrating the California Standards Test in Mathematics into the API*, March 2002 at 26; Advisory Committee, *The 2002 API: Integrating the CAHSEE Results into the API*, May 2002; SBE Minutes & Agenda Item #5, January 8, 2003.

CHART 9
API Six-Year Plan†



† Adapted from: CDE, *The API: A Six-Year Plan for Development (2001-2006)*, April 25, 2002 at 12.

* Pending Board adoption.

CHART 10
Summary of API Award Programs*

Characteristic	Governor's Performance Awards (GPA) [Accountability Act]	Certificated Staff Performance Incentive Act [AB 1114]	School Site Employee Performance Bonus [SB 1667]
	2000 = \$227 million 2001 = 144.3 million	\$100 million	\$350 million proposed
	School	School Teachers & Principals	All staff at school site <i>plus</i> the school
	Open to All Schools with APIs	Open to schools with APIs in Deciles 1-5 in 1999	Open to all schools with APIs
	<ul style="list-style-type: none"> ✓ API 5% Growth Target ✓ Comparable Improvement 4% ✓ Participation Rate = 95% ✓ If API = 800+, gain ≥ 1 point 	<ul style="list-style-type: none"> ✓ 1998-99 Stanford Test growth ✓ API Growth at least 10% ✓ Comparable Improvement 8% ✓ Participation Rate = 95% 	<ul style="list-style-type: none"> ✓ Eligibility for GPA program will determine eligibility for the performance bonus
	<ul style="list-style-type: none"> ✓ Intended to be fully funded at \$150 per student to all schools meeting conditions ✓ Not fully funded so alternate distribution system used: <ul style="list-style-type: none"> • Biggest API gains received the most \$ • 20% of funds distributed at \$150 per student • 30% at \$100 per student • 50% at \$50 per student • Some eligible schools received non-monetary rewards only • Within award levels, schools proportionately distributed by school level (Elem, Mid, HS) 	<ul style="list-style-type: none"> ✓ Biggest API gains receive the most \$ <ul style="list-style-type: none"> • 1000 teachers in schools with largest API growth get \$25,000 each • 3750 teachers get \$10,000 each • 7500 teachers get \$5,000 each 	<ul style="list-style-type: none"> ✓ All site staff (on FTE basis) will receive bonus ✓ An equal amount of money given to the school for school-wide use
	Site governance team; Ratified by local board	Local district working with teachers' union	Same group as the one used to make decisions for GPA program
	NOT funded after 2001 due to budget deficit	NOT funded after 2000	One-time bonus

* Adapted from: SBE, Minutes & Agenda Item #11, July 12, 2000, Attachment 6; updated from: www.cde.ca.gov/psaa/awards.



Coronado Elementary School

2001 Virginia Ave., Richmond, CA 94804 PHONE: (510) 233-7800
PRINCIPAL: Linda Jackson GRADE LEVELS: K-5 SCHEDULE: Normal

Principal's Comments

Coronado School provides students with the highest quality education possible and helps students make positive life choices, strengthen our community, and successfully participate in a diverse and global society. We provide excellent learning and teaching experiences, a safe student-centered learning environment, and support for both students and staff.

As an Immediate Intervention Underperforming School Program (Tide I, II/USP) school, Coronado's entire staff receives intense training in curriculum aligned with state standards. Coronado has gained 162 API points over a two-year period, ranking us 24th in the state of California and number one in the Oakland-San Francisco Bay Area.

Coronado School has implemented the 21st Century After-school Tutorial and Enrichment Program, extended library hours, and reduced class sizes. We have also implemented parenting classes for both English and Spanish speakers, as well as a summer academy for students.

Teachers and Students

KEY FACTOR	OUR SCHOOL	COUNTY AVG	STATE AVG
Students	415	532	582
Teachers	23	28	31
Students per teacher	18	19	19

Teachers, 2000-2001

Teachers have varying levels of experience and credentials. The teacher shortage has required almost all schools to hire some teachers lacking full credentials.

KEY FACTOR	OUR SCHOOL	COUNTY AVG	STATE AVG
Average years teaching experience	12	13	12
Male teachers	17%	11%	16%
Elementary credential holders	100%	93%	92%
Trainee credential holders	9%	3%	2%
Emergency permit holders	18%	8%	10%

Average Class Sizes, 2000-2001

GRADE	OUR SCHOOL	COUNTY AVG	STATE AVG
Kindergarten	18	21	19
First grade	19	19	19
Second grade	15	19	19
Third grade	20	19	19
Fourth grade	23	28	29
Fifth grade	32	29	29

Students, 2000-2001

Students bring different literacy skills to school. We provide information on the three factors below because they may affect how well students perform in school.

KEY FACTOR	OUR SCHOOL	COUNTY AVG	STATE AVG
Students still learning English	31%	29%	32%
Students qualifying for free/reduced-price lunch	97%	34%	54%
Students whose parents attended/graduated college	13%	65%	55%

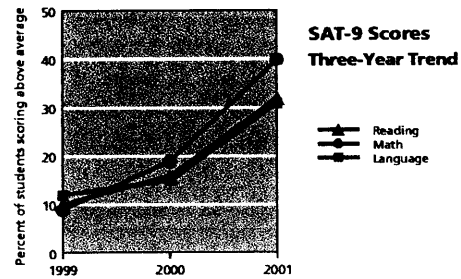
How Students Are Achieving

Academic Performance Index, Spring 2001

The Academic Performance Index (API) is a way of comparing schools based on their students' test scores. Scores are based on a scale from 200 to 1,000.

YEAR 2001 API	GROWTH ATTAINED	MET TARGET GROWTH	BONUS AWARDED	UNDERPERFORMING SCHOOL
543	+109	Yes	Yes	Yes

Stanford-9 (SAT-9) Test, Spring 2001

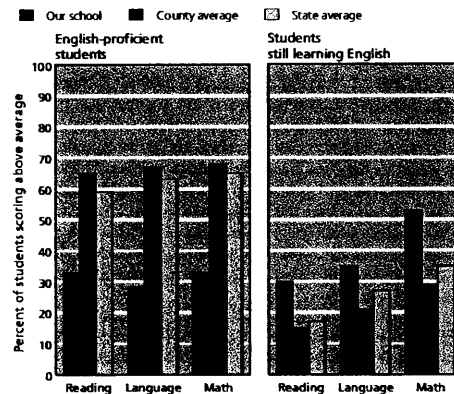


Almost all students in grades two through five took the Stanford-9 (SAT-9) test in the spring of 2001. The resulting scores measure how well students performed compared to other students. The results below indicate what percent of students scored above average (50th percentile or higher).

SUBJECT	PERCENT OF STUDENTS SCORING ABOVE AVERAGE		
	OUR SCHOOL	COUNTY AVG	STATE AVG
Reading	32%	59%	48%
Language	31%	61%	53%
Math	40%	63%	57%

The test scores of students who are English-proficient, and those who are still learning English differ dramatically. Below you'll see the scores of each subgroup separated.

SAT-9 Scores By English-language proficiency



Resources

TEXTBOOKS: We use these textbooks for the core curricula:

TITLE	SUBJECT AREA	YEAR OF PUBLICATION	ENOUGH FOR EACH STUDENT?
Reading/Mastery, SRA	Reading	1999	Yes
Harcourt Brace	Math	2000	Yes
FOSS Kits	Science	1992	N/A
Harcourt Brace	Social studies	2000	Yes

LIBRARY: The Library/Media Center is equipped with three computers, one printer, and a scanner. As of May 2001, the school library contained 6,083 volumes of books, ranging from easy readers to reference books. There are two part-time librarian assistants, allowing for hours to be extended to 6 p.m. Monday through Friday.

COMPUTERS: Our Computer Lab is equipped with 24 Pentium computers. Classes visit the lab once weekly and receive instruction in technology from a part-time technician. The lab is also used three days a week during our afterschool program. Follow-up lessons are taught in the computer lab or in their classrooms.

KEY FACTOR	OUR SCHOOL	COUNTY AVG	STATE AVG
Students per computer	8	6	7
Internet-connected classrooms	13	27	18

BUILDINGS: Coronado School has a comprehensive school safety plan, which includes appropriate strategies and programs that provide a high level of safety for students. Also included are procedures of reporting and preventing crime, reporting child abuse, designing disaster procedures and developing policies for suspension and expulsion. Additionally, the plan contains the sexual harassment policy, school-wide dress code, procedures for the safe coming and going of children, parents, and employees, and rules and procedures for school-wide discipline. Staff members strive to provide a safe and orderly environment conducive to learning.

Climate for Learning

HOMEWORK: Coronado students are expected to complete up to one hour of homework assignments four days a week. Some teachers present homework packets that are turned in weekly, while other teachers distribute homework nightly. Homework is expected to be presented neatly and on time. Homework also includes 20 minutes of reading at home.

ATTENDANCE: The attendance rates of students are indicators of how connected they are to the school.

KEY FACTOR	OUR SCHOOL	DISTRICT AVG
Students	92%	94%

DISCIPLINE: A clear explanation of Coronado Cougar Rules is given to parents and students at the beginning of each school year. These rules prohibit behaviors and activities that interfere with instruction, learning, and achievement. Parents are asked to discuss Coronado Cougar Rules with their children. Students are expected to follow these guidelines for their own and others' safety and well-being. These rules support our philosophy that all students should have a safe and positive school experience. The principal discusses these rules with students in grades one through three at the daily morning assembly.

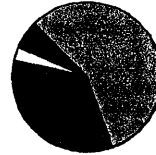
Note that suspensions and expulsions reflect both how strict our rules are, and how strictly we enforce them. We do not count in-school suspensions.

DISCIPLINE FACTOR	OUR SCHOOL	DISTRICT AVG
Suspensions per 100 students	4	11
Expulsions per 100 students	0	0

SAFETY: Coronado School has a comprehensive school safety plan, which includes appropriate strategies and programs that provide a high level of safety for students. Also included are procedures of reporting and preventing crime, reporting child abuse, designing disaster procedures and developing policies for suspension and expulsion. Additionally, the plan contains the sexual harassment policy, school-wide dress code, procedures for the safe coming and going of children, parents, and employees, and rules and procedures for school-wide discipline. Staff members strive to provide a safe and orderly environment conducive to learning.

How We Spend Our Time

Time spent in each subject differs at each grade level. We use fifth grade as a typical example.



Fifth-Grade Classroom Time
Portion of time spent per subject

Reading/writing
 Math
 Science
 Social studies

SCHEDULE: Our school year consists of 181 days. School starts at 8:15 a.m. for all students. Students in primary grades begin the morning with a 45-minute assembly with the principal. There is an afterschool program five days a week which ends at 6 p.m. Monday through Friday.

Specialized Programs

Some students are in specialized programs for some part of the day. The percent of students in each program follows.

PROGRAM	OUR SCHOOL	DISTRICT AVG
English as a Second Language	31%	27%
Gifted/Talented (GATE)	0%	2%
Special Education	10%	11%

Major Achievements

- Increased the number of students performing in the second and third quartile in reading and mathematics.
- Received state Academic Performance awards for two consecutive years.
- Had a record number of parents participate in parent workshops and Parent Night activities.

Focus For Improvement

- Increase the amount of fluent readers by the end of each grade level.
- Increase student attendance to 97%.
- Provide afterschool reading and writing in Spanish for students in grades four and five.

Keeping You Informed

This report provides information about how well our school is doing—where it is succeeding and where there is room for improvement. While it cannot tell you everything about our results, the report is a good starting point for discussions with teachers and principals. For more information, to see a full-length version of this report, or to access the technical data appendix to this report, please contact the school.

Coronado Elementary School
2001 Virginia Ave.
Richmond, CA 94804
Phone: (510) 233-7800

West Contra Costa USD
1108 Bissell Avenue
Richmond, CA 94801
(510) 234-3825

NOTES ON SOURCES: Student and teacher data, 2000–2001 school year. SAT-9 data, spring 2001. API, October 2001.



PUBLISHED BY SCHOOL WISE PRESS

www.schoolwisepress.com © 2001 by Publishing 20/20

C H A R T 1 1 b
2002 API Growth Report for Coronado Elementary School

California Department of Education
 Policy and Evaluation Division

**2001-2002 Academic Performance Index (API) Growth Report
 School Report**

Revised April 3, 2003

School: Coronado Elementary
 District: West Contra Costa Unified
 County: Contra Costa
 CDS Code: 07 -61796 -6004667

School Type: Elementary

STAR 2002 Percent Tested	100	Number of Students Included in the 2002 API (Growth)	244	2002 API (Growth)	591	2001- 2002 Growth Target	13	2001- 2002 Growth	50	<u>Met Growth Target</u>		Awards Eligible
										Comparable Improve- ment (CI)	Both Schoolwide and CI	Yes
										Yes	Yes	Yes

"N/A" means a number is not applicable or not available due to missing data.
 "*" means this API is calculated for a small school, defined as having between 11 and 99 valid STAR test scores. The API is asterisked if the school was small either in 2001 or 2002. APIs base on small numbers of students are less reliable and therefore should be interpreted with caution.
 "A" means the school scored at or above the interim Statewide Performance Target of 800 in 2001.
 "B" means the school is not awards eligible to Adult Testing Irregularities.

Awards Notes - The "Awards Eligible" column requires at least five points schoolwide growth and at least four points for growth for each numerically significant

For more details about the displayed information, see the [Explanatory Notes for the 2001-2002 API \(Growth\) Report](#).

Subgroups

	Number of Pupils Included in 2002 API	Numerically Significant in both Years	2002 Subgroup API (Growth)	2001 Subgroup API (Base)	2001-2002 Subgroup Growth		Met Subgroup Growth Target
					Target	Growth	
Ethnic/Racial							
African American (not of Hispanic origin)	129	Yes	566	500	10	66	Yes
American Indian or Alaska Native	1	No					
Asian	1	No					
Filipino	2	No					
Hispanic or Latino	105	Yes	621	605	10	16	Yes
Pacific Islander	2	No					
White (not of Hispanic origin)	4	No					
Socioeconomically Disadvantaged	242	Yes	593	535	10	58	Yes

"A" means the subgroup scored at or above the interim Statewide Performance Target of 800 in 2001.

School Demographic Characteristics

These data are from the October 2001 California Basic Educational Data System (CBEDS) data collection, the 2002 Standardized Testing and Reporting (STAR) student answer document, and the STAR Apportionment data collection.

Ethnic/Racial (STAR)	Percent	Parent Education Level (STAR)	Percent
African American (not of Hispanic origin)	52	Percent with a response*	88
American Indian or Alaska Native	0	Of those with a response:	
Asian	0	Not a high school graduate	38
Filipino	1	High school graduate	53
Hispanic or Latino	44	Some college	8
Pacific Islander	1	College graduate	2
White (not of Hispanic origin)	1	Graduate school	0

These percentages may not sum to 100 due to responses of: other, multiple, declined to state, or non-response.

*This number is the percentage of student answer documents with stated parent education level information.

Participants in Free or Reduced Price Lunch (STAR)	99	Average Parent Education Level (STAR)	<u>Average</u> 1.73
English Learners (STAR)	39	<i>The average of all responses where "1" represents "Not a high school graduate" and "5" represents "Graduate school."</i>	
Multi-track year-round school (CBEDS)	No	Fully credentialed teachers (CBEDS)	<u>Percent</u> 74
School Mobility (STAR)	20	Teachers w/emergency credentials (CBEDS)	11
<i>This is the percent of students who first attended this school in the current year.</i>		Enrollment in grades 2-11 on first day of testing (STAR Apportionment)	<u>Number</u> 269
District Mobility (STAR)	9	Number of students excused from STAR testing (STAR)	0
<i>This is the percent of students who first attended this district in the current year.</i>		<i>Students required to have alternative assessments due to IEP exemptions</i>	
Average Class Size (CBEDS)		Number of Students Tested (STAR)	269
Grades	<u>Average</u>	<i>Students excused per parent written request</i>	
K-3	18		
4-6	21		
Core academic courses in departmentalized programs.	N/A		

CHART 12a

Sample Stanford Test & Standards Tests Student Report



STANDARDIZED TESTING AND REPORTING (STAR) PROGRAM

Reporting 2002 STAR Results to Parents/Guardians

Sample Front Page 1 of Performance Report – Grade 9

Standardized Testing and Reporting (STAR) Performance Report

This is a report to explain your child's academic performance on a state test he or she took this past spring. It is divided into two parts.

The first part, which begins below, tells you how your child performed in meeting California's academic standards. These standards make clear what all students are required to learn at each grade level. An explanation of these requirements begins on the back of this page.

The second part, which is on the next page, tells you how your child's test results compare to those of other students across the United States.

The two sections combined should help you understand how your child is doing in school. You can get additional information about these test results from your child's teacher. Information about the tests and standards is available on the Internet at www.cde.ca.gov/statetests/star.

Report for

Bianca H Mata

Student No. 000

DOB: 02/12/88 Grade: 9 Test Date: 05/02

Teacher: Michaelson (0000789012)

School: Johnson Middle Sch (0009544)

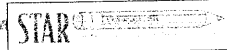
District: Langeberg Unified (3456789)

Parents of

Bianca H Mata

123 Main Street

Los Angeles, California 90210



Academic Standards: California Standards Test – Grade 9

This report indicates your child's performance on test questions that reflect California's standards of what a student should know and be able to do at each grade level. There are separate standards for English-language arts, mathematics, history-social science, and science. In grades 2-8, students are tested in English-language arts and mathematics only.

The *overall results* show your child's overall score for each subject and whether he or she is exceeding, meeting or falling below the standards. The *specific results* show how your child performed on specific components of the standards.

English-Language Arts

Overall Results				State Targets for All Students	
Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced
359				◆	

Your child's performance level is based on his or her overall score. In English-language arts, scores are:

- Far Below Basic: a score below 264
- Below Basic: 265-299
- Basic: 300-349
- Proficient: 350-396
- Advanced: 397 or higher

Specific Results		
English-Language Arts Components	Total Questions	Number Correct
Reading	50	36
Word Analysis and Vocabulary	10	7
Reading Comprehension	21	18
Literary Response and Analysis	19	11
Writing	40	31
Writing Strategies	16	11
Written and Oral Language Conventions	24	20

Mathematics: Geometry

Overall Results				State Targets for All Students	
Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced
383				◆	

Your child's performance level is based on his or her overall score. In Geometry, scores are:

- Far Below Basic: a score below 246
- Below Basic: 247-299
- Basic: 300-349
- Proficient: 350-417
- Advanced: 418 or higher

Specific Results		
Geometry Components	Total Questions	Number Correct
Logic and Geometric Proofs	23	18
Volume and Area Formulas	11	7
Angle Relationships, Constructions, and Lines	16	14
Trigonometry	15	12



Source: www.cde.ca.gov/statetests/star



Sample Front Page 2 of Performance Report – Grade 9

California Standards Test – Grade 9, continued

Report for Bianca H Mata

History-Social Science Cumulative

Overall Results				State Targets for All Students	
Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced
503				◆	

Your child's performance level is based on his or her overall score. In history-social science cumulative, scores are:

- Far Below Basic: a score below 270
- Below Basic: 271-299
- Basic: 300-349
- Proficient: 350-395
- Advanced: 396 or higher

Specific Results		
History-Social Science Components	Total Questions	Number Correct
California: A Changing State; and U.S. History and Geography; Making a New Nation	15	14
World History and Geography: Ancient Civilizations	11	11
World History and Geography: Medieval and Early Modern Times	14	13
U.S. History and Geography: Growth and Conflict	20	18

Sciences: Earth Science

Overall Results				State Targets for All Students	
Score	Far Below Basic	Below Basic	Basic	Proficient	Advanced
383				◆	

Your child's performance level is based on his or her overall score. In Earth Science, scores are:

- Far Below Basic: a score below 276
- Below Basic: 277-299
- Basic: 300-349
- Proficient: 350-392
- Advanced: 393 or higher

Specific Results		
Earth Science Components	Total Questions	Number Correct
Investigation and Experimentation	9	6
Astronomy and Cosmology	16	13
Solid Earth	12	8
The Earth's Energy	23	19

National Comparison: Stanford Achievement Test Series, Ninth Edition - grade 9 test

This part of the report compares your child's performance with that of children across the country. Your child's score is reported as a percentile. The higher the score, the better your child's ranking on the test. For example, a student who scores in the 40th percentile performed as well as 40 percent of all students nationally - but not as well as 60 percent. A student who scores in the 90th percentile performed as well as 90 percent of all students. The table below displays your child's score for each area tested, including the number of questions on the test, the number your child answered correctly, and his or her national percentile rank.

Subtests and Totals	Total Questions	Number Correct	Non-Standard	Student's Percentile Rank				
				Below Average	Average	Above Average	1	99
Reading	84	58		10	49	70	90	99
Vocabulary	30	19			43			
Reading Comprehension	54	39			53			
Mathematics	48	19			37			
Language	48	25			31			
Language Mechanics	24	9			17			
Language Expression	24	16			48			
Science	40	22			64			
Social Science	40	17			48			

California Reading List Number

Your child's reading list number is **12**

You can use this number to get a list of state-recommended books that are at your child's reading level based on his or her Stanford 9 Reading Comprehension Score. For a copy of the reading list, visit the STAR web site (<http://star.cde.ca.gov>).

1995 NORMS - Spring, National

Scores based on normative data, 1996, by Harcourt, Inc. All rights reserved.





Sample School Report – Mathematics Demographic Summary for All Students Tested

California High School Exit Examination
 Demographic Summary for All Students Tested
 Mathematics (March 2002)

- [To Research Files](#)
- [CAHSEE web site](#)
- [DataQuest Home Page](#)

County: 00 – EXAMPLE COUNTY
 District: 00000 – EXAMPLE SCHOOL DISTRICT
 School: 0000000 – EXAMPLE HIGH SCHOOL

Number of Students Tested: 452
 Number of Students Enrolled: 840

	Number Tested	Number Passed	Percent Passed	Number Not Passed	Percent Not Passed	Mean Scaled Score	Strands for Mathematics (Average Percent Correct)				
							Probability & Stat.	Number Sense	Algebra & Func.	Meas. & Geometry	Algebra I
All Students Tested (Average)	452	271	60%	181	40%	375	92%	68%	53%	61%	92%
Grade											
Tenth	452	271	60%	181	40%	375	92%	68%	53%	61%	92%
Eleventh	0	–	–%	–	–%	–	–%	–%	–%	–%	–%
Twelfth	0	–	–%	–	–%	–	–%	–%	–%	–%	–%
Adult Ed.	0	–	–%	–	–%	–	–%	–%	–%	–%	–%
Unknown	0	–	–%	–	–%	–	–%	–%	–%	–%	–%
Gender											
Female	217	130	60%	87	40%	375	92%	68%	53%	61%	92%
Male	210	126	60%	84	40%	375	92%	68%	53%	61%	92%
Unknown	25	15	60%	10	40%	375	92%	68%	53%	61%	92%
Ethnicity											
American Indian or Alaskan Native	5	–	–%	–	–%	–	–%	–%	–%	–%	–%
Asian	108	65	60%	43	40%	375	92%	68%	53%	61%	92%
African-American	68	41	60%	27	40%	375	92%	68%	53%	61%	92%
Filipino	50	30	60%	20	40%	375	92%	68%	53%	61%	92%
Hispanic or Latino	104	62	60%	42	40%	375	92%	68%	53%	61%	92%
Pacific Islander	9	–	–%	–	–%	–	–%	–%	–%	–%	–%
White (not of Hispanic origin)	83	50	60%	33	40%	375	92%	68%	53%	61%	92%
Unknown	25	15	60%	10	40%	375	92%	68%	53%	61%	92%
Language Fluency											
English Learner (EL)	35	21	60%	14	40%	375	92%	68%	53%	61%	92%
Initially Fluent English Proficient (IFEP)	25	15	60%	10	40%	375	92%	68%	53%	61%	92%
Redesignated Fluent English Proficient (RFEP)	45	27	60%	18	40%	375	92%	68%	53%	61%	92%
English Only	322	193	60%	129	40%	375	92%	68%	53%	61%	92%
Unknown	25	15	60%	10	40%	375	92%	68%	53%	61%	92%
Economic Status											
Economically Disadvantaged Students	185	111	60%	74	40%	375	92%	68%	53%	61%	92%
Non-Economically Disadvantaged Students	212	127	60%	85	40%	375	92%	68%	53%	61%	92%
Unknown	55	33	60%	22	40%	375	92%	68%	53%	61%	92%
Special Education Program Participation											
Students Receiving Services	8	–	–%	–	–%	–	–%	–%	–%	–%	–%
Students Not Receiving Services	444	266	60%	178	40%	375	92%	68%	53%	61%	92%

– To protect privacy, no results for any group with fewer than 11 students will be released.

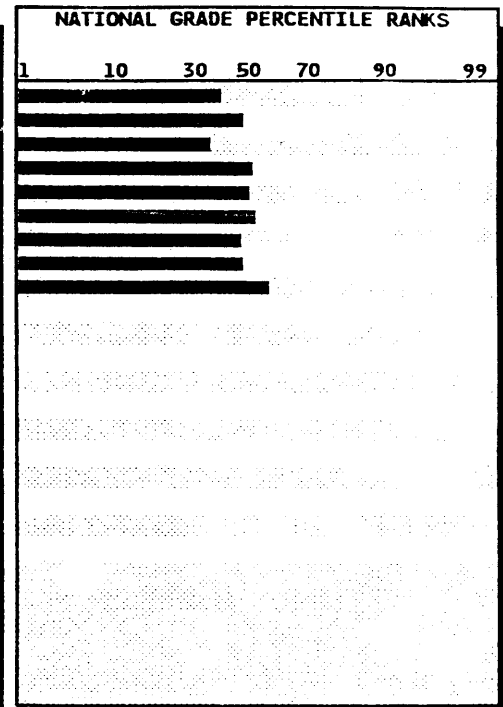
You may obtain copies of selected test questions at your school site or at the following Web site: <http://www.cde.ca.gov/statetests/cahsee>.

GROUP REPORT FOR CALIFORNIA

GRADE: 11
TEST DATE: 05/02

TEST TYPE: MULTIPLE CHOICE

<i>SUBTESTS AND TOTALS</i>	Number Tested	Mean Raw Score	Mean Scaled Score	National Indiv PR-S	Mean National NCE
Total Reading	316809	54.3	699	38-4	43.6
Vocabulary	318851	19.5	718	47-5	48.5
Reading Comp.	320625	34.6	689	34-4	41.4
Mathematics	317382	22.8	705	51-5	50.6
Language	316254	28.9	684	49-5	49.4
Lang Mechanics	316234	14.2	687	52-5	51.1
Lang Expression	316216	14.7	682	46-5	47.7
Science	316000	18.2	685	47-5	48.3
Social Science	316121	18.1	668	57-5	53.8



<i>CONTENT CLUSTERS</i>	Number of Items	PERCENT IN EACH		
		Below Average	Average	Above Average
Reading Vocabulary	30	23	56	21
Synonyms	16	23	50	27
Context	7	26	48	26
Multiple Meanings	7	31	57	13
Reading Comprehension	54	35	51	14
Recreational	18	32	40	28
Textual	18	34	47	19
Functional	18	35	48	17
Initial Understanding	10	33	45	23
Interpretation	24	29	51	20
Critical Analysis	10	33	47	20
Process Strategies	10	41	29	30
Mathematics	48	26	47	27
Problem-Solving Strategies	6	13	51	36
Algebra	6	23	60	17
Statistics	6	18	51	31
Probability	5	17	50	33
Functions	5	23	43	34
Geometry from a Synthetic Perspective	6	22	51	27
Geometry from an Algebraic Perspective	5	38	39	22
Trigonometry	3	23	56	21
Discrete Mathematics	3	20	66	14
Conceptual Underpinnings of Calculus	3	26	37	37
Language	48	24	52	24
Capitalization	8	27	57	16
Punctuation	8	24	47	29
Usage	8	13	57	30
Sentence Structure	12	18	63	19
Content and Organization	12	28	49	23

<i>CONTENT CLUSTERS</i>	Number of Items	PERCENT IN EACH		
		Below Average	Average	Above Average
Science	40	30	47	23
Earth & Space Science	13	27	40	32
Physical Science	14	34	44	22
Life Science	13	34	49	17
Science Process Skills	32	41	39	20
Social Science	40	18	51	31
History	10	14	45	41
Geography	9	21	60	19
Civics & Government	8	16	56	28
Economics	8	14	73	13
Culture	5	30	25	45

CHART 12c

1998-2001 Intervention Program Schools' APIs & Test Scores

American Institutes for Research

Exhibit 3.2: API, SAT-9 Math, and SAT-9 Reading Scores: Estimated Average Achievement for Cohort 1 II/USP and Comparison Elementary Schools in 1998, 1999, 2000, and 2001

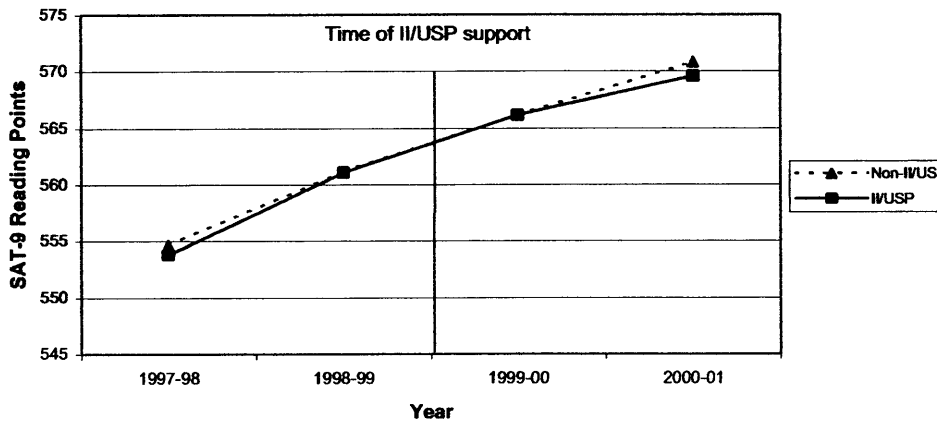
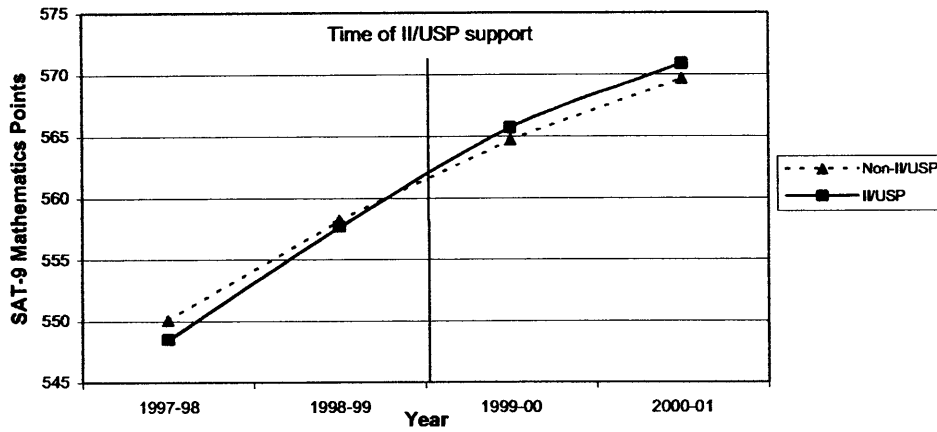
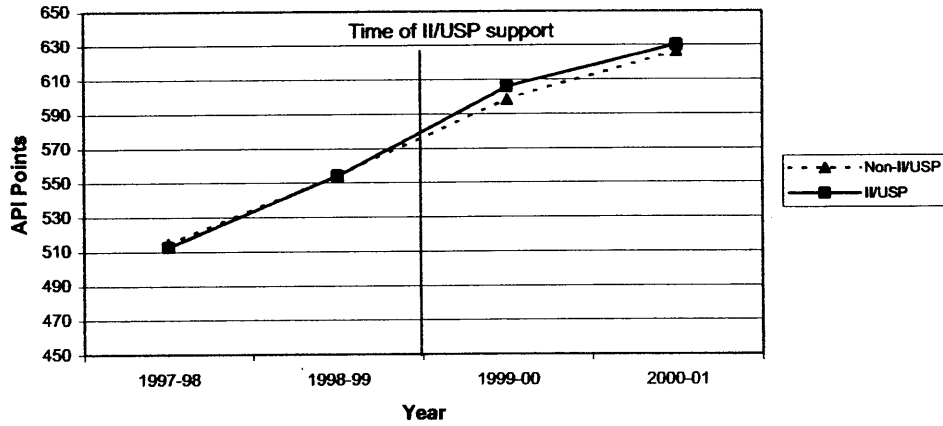
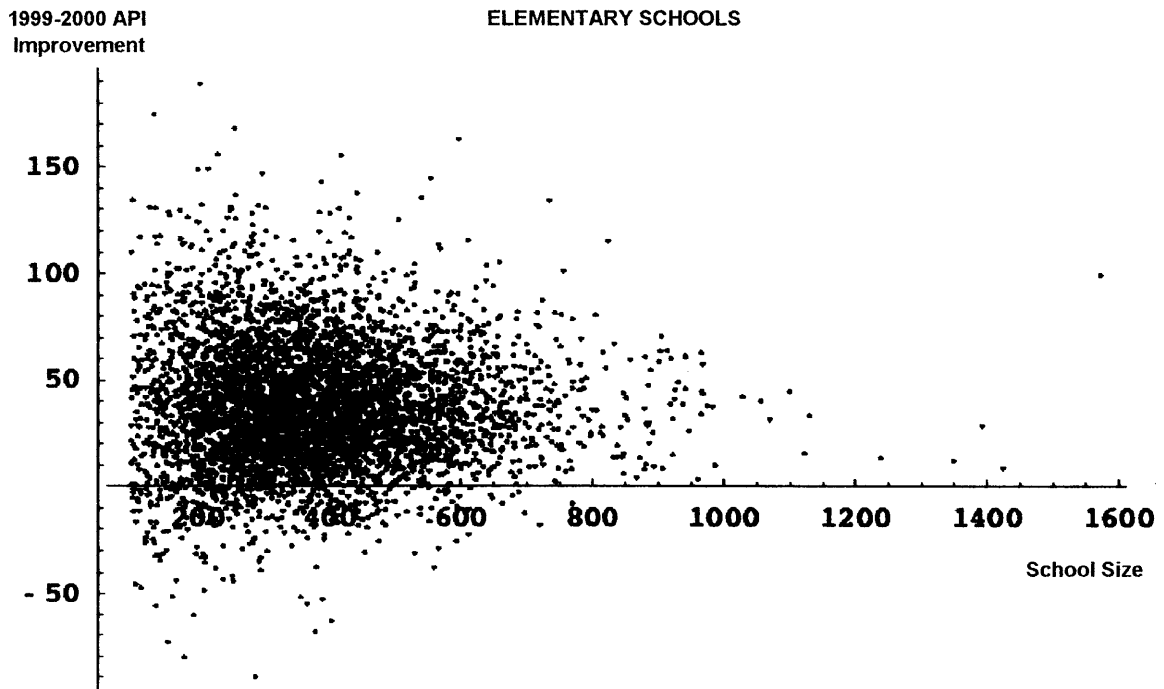
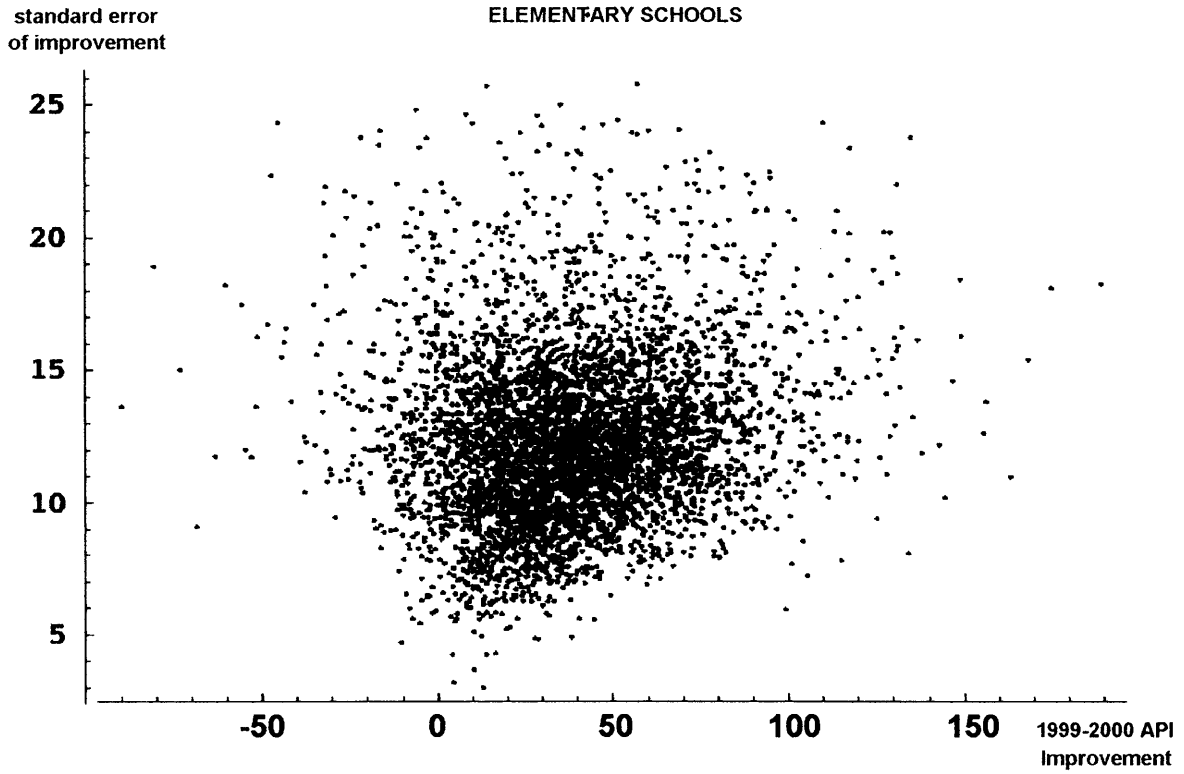
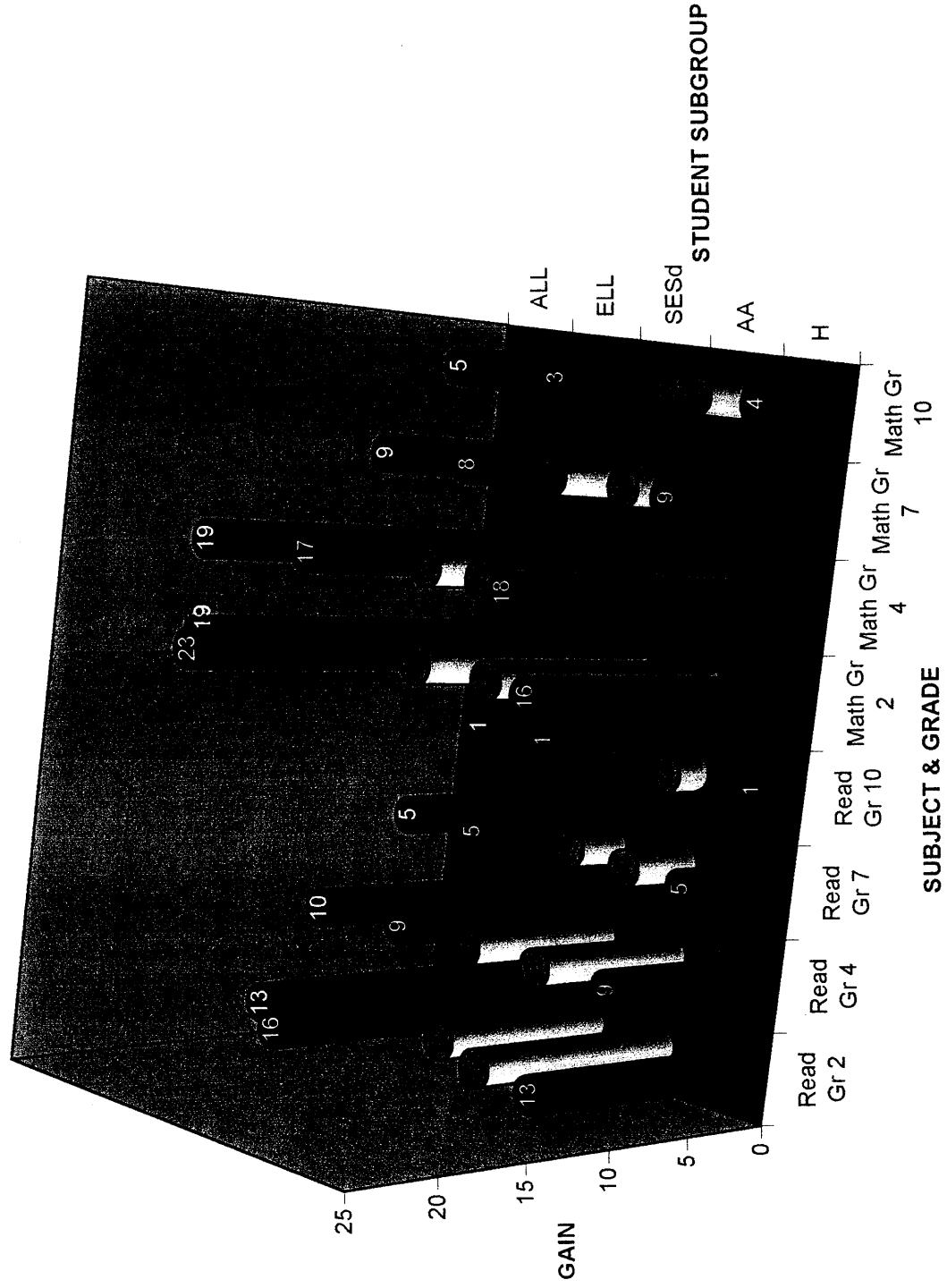


CHART 13
Relationship Between API Improvement, School Size and Standard Error*



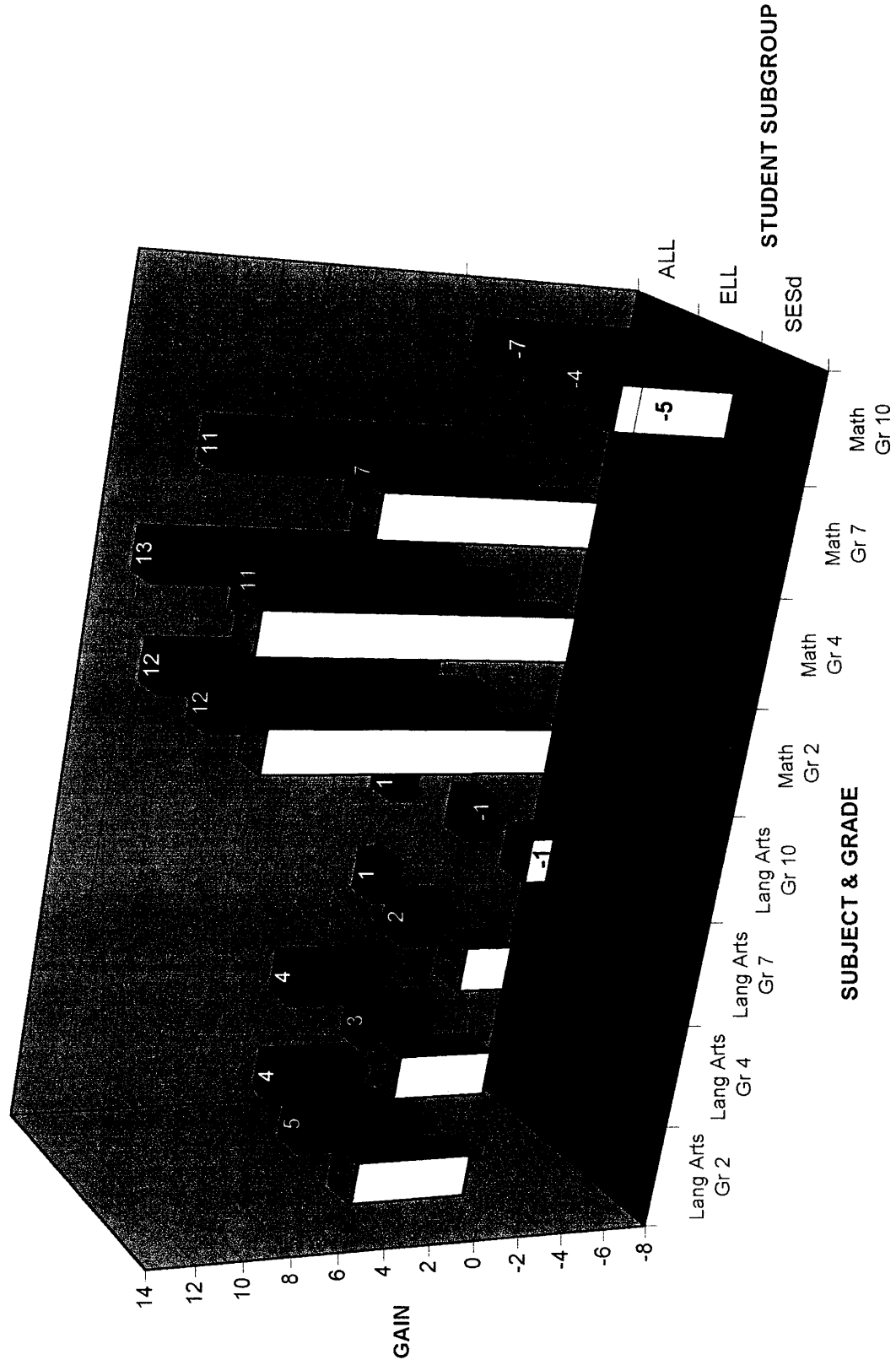
* Source: Rogosa, D., *Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger "Volatility in School Test Scores"*, Stanford University, October 2002, www.api.cde.ca.gov, Figure 2.1, 2.2 (bottom), p. 34, 36.

CHART 14a
 2002-1998 NPR Gains in "Average" Stanford Test Scores*



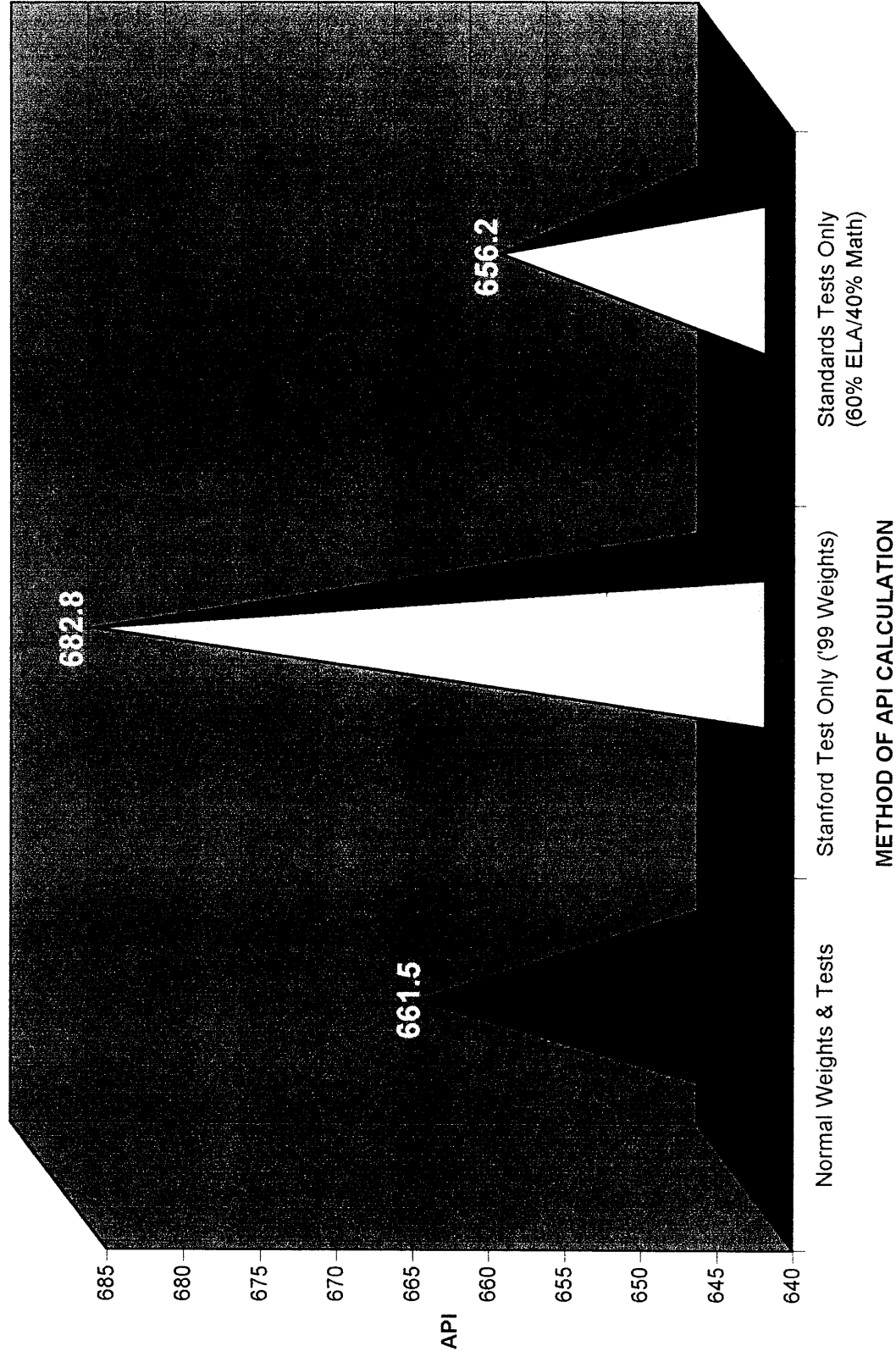
* Gains in NPR for "Average" Student Score for ALL, ELL, SESd;
 Gains in Percent Above 50th NPR for AA and H;
 SESd, AA and H 2002-1999.

CHART 14b
2001-1999 Gains in Average Percent Correct for Standards Tests*



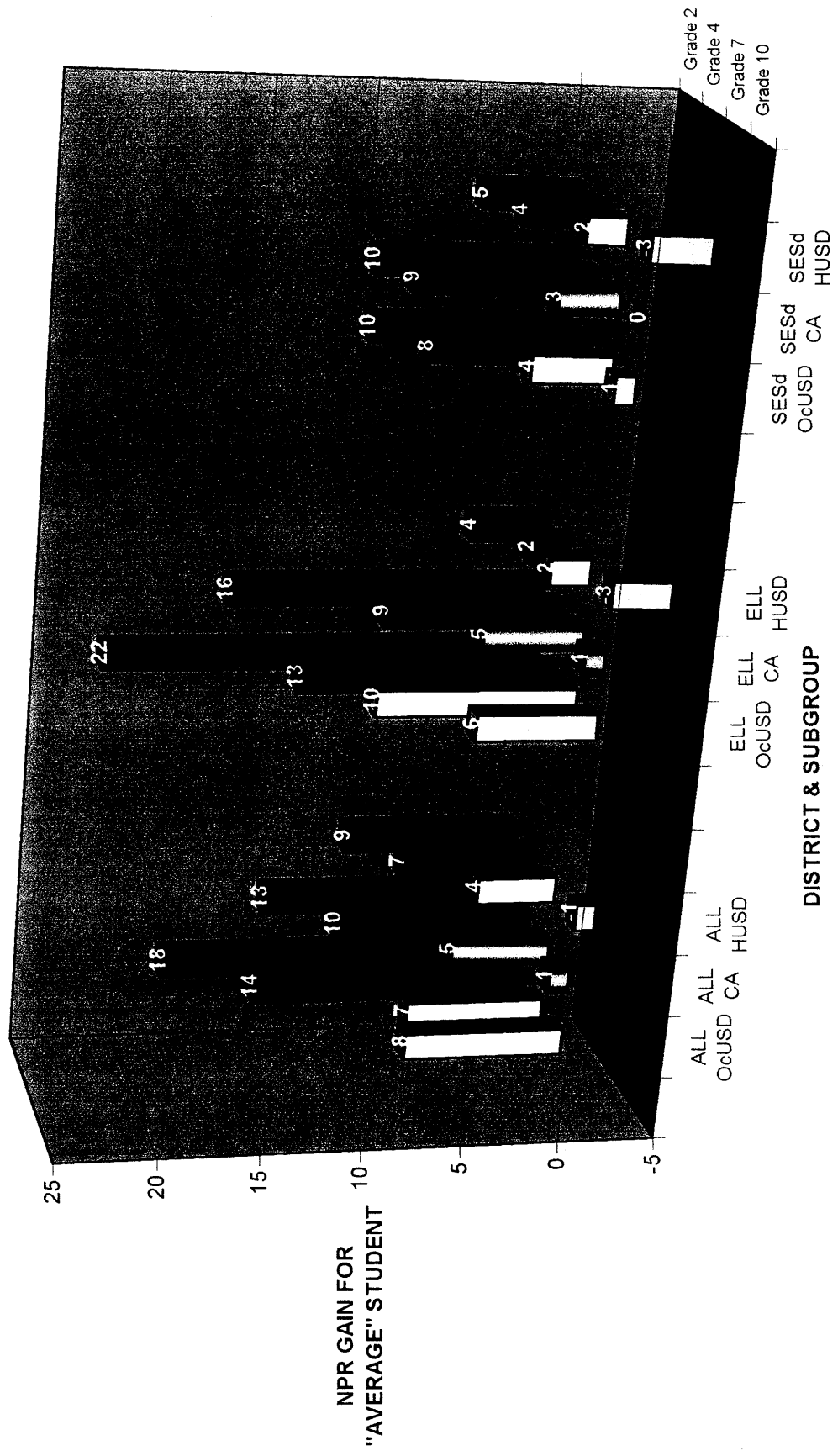
* Not equated across years

CHART 15
2002 State Base Stanford Test Only versus Standards Tests Only APIs



* Statewide, Grades 2-8 Only, Non-mobile Students

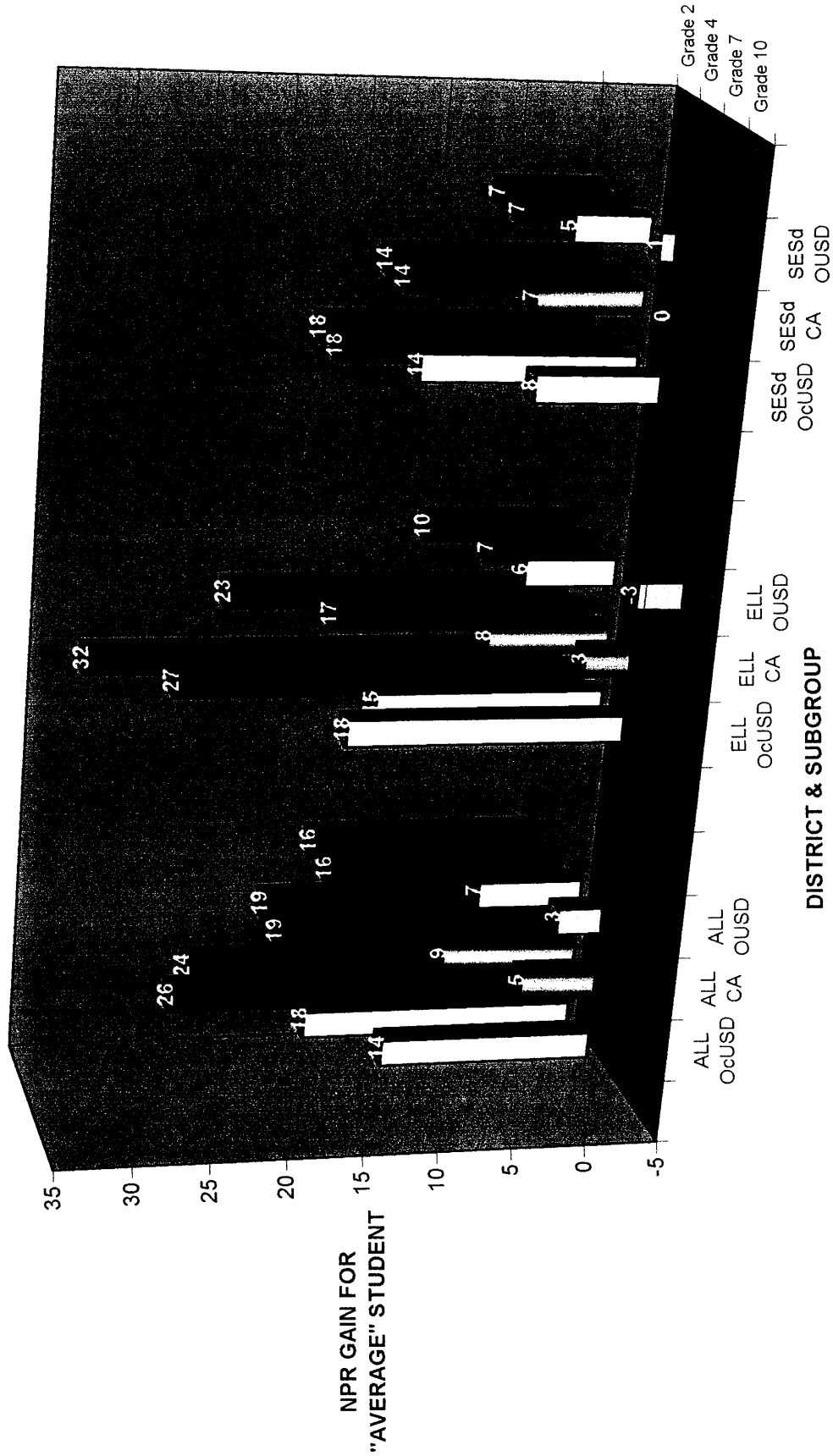
CHART 16a
2002-1998 Stanford Test Reading Growth by District for Selected Grades*



* 1999 data for SESs.

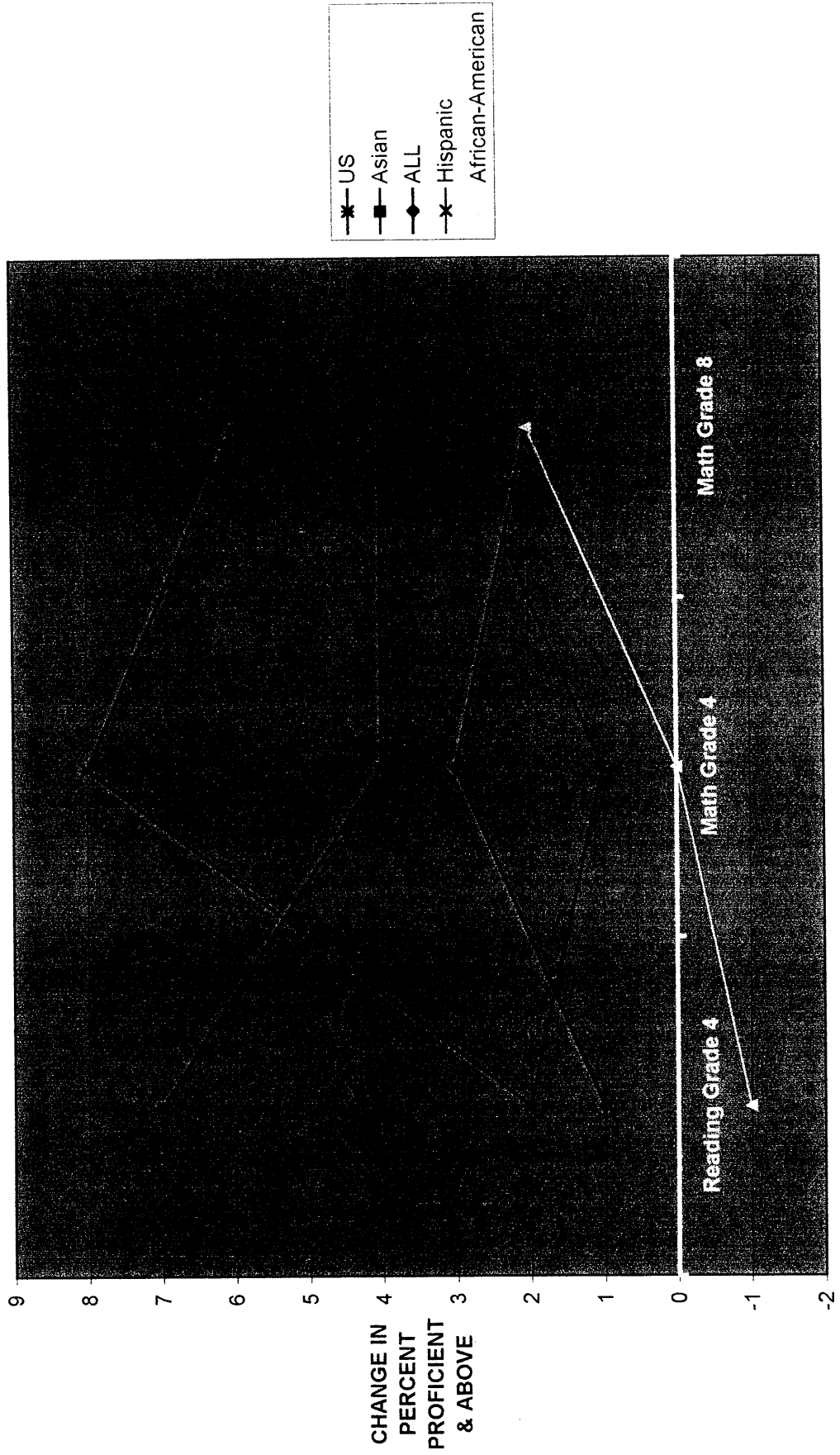
CHART 16b

2002-1998 Stanford Test Mathematics Growth by District for Selected Grades*



* 1999 data for SES.

CHART 17
 California Subgroup & US NAEP Changes in Percent Proficient & Above



* Reading 1988-1992
 Mathematics 2000-1992

CHART 18
2000-2002 API Awards Eligibility by Decile

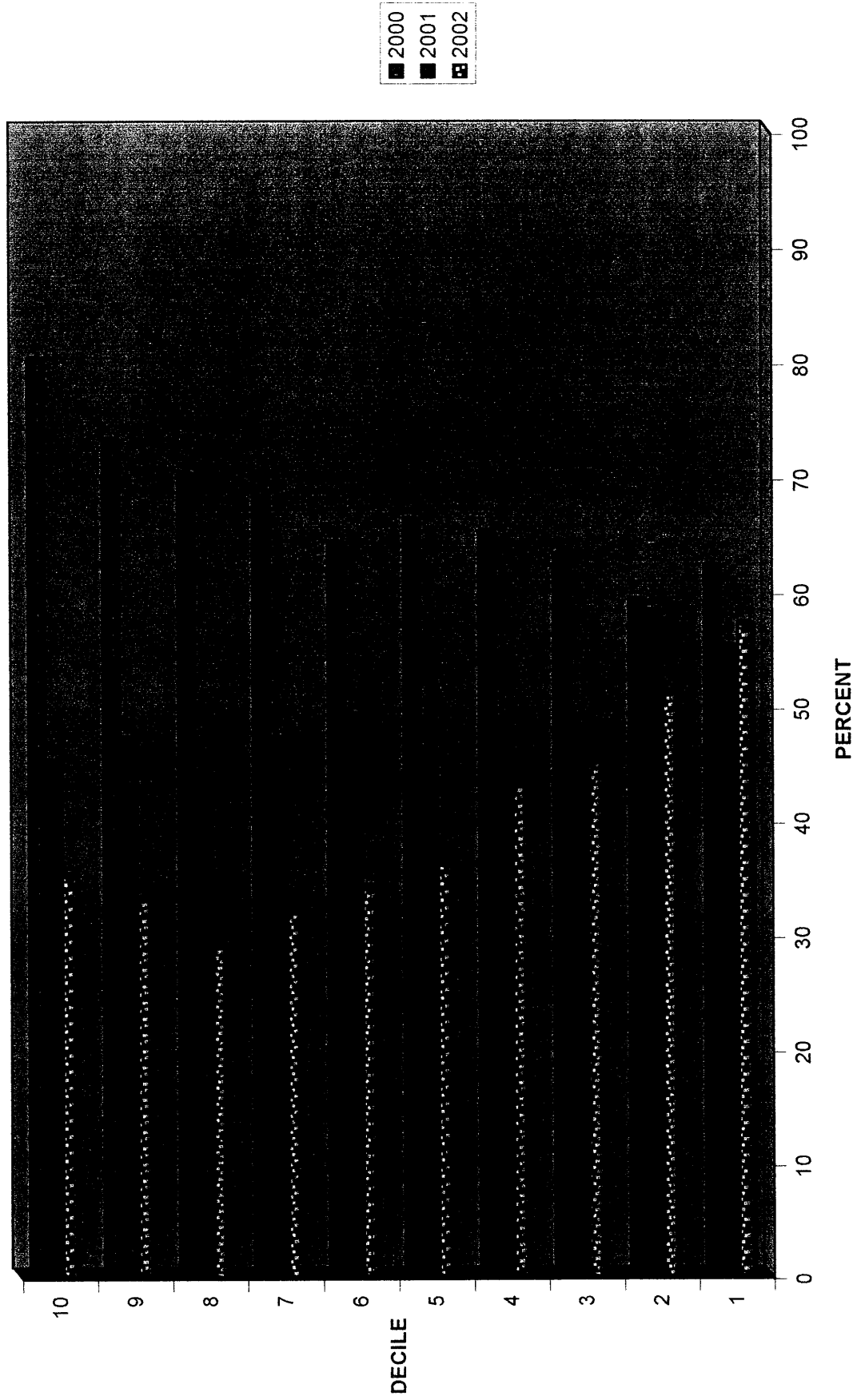


CHART 19
Awards & Intervention Funding for Schools Attended by Named Plaintiffs

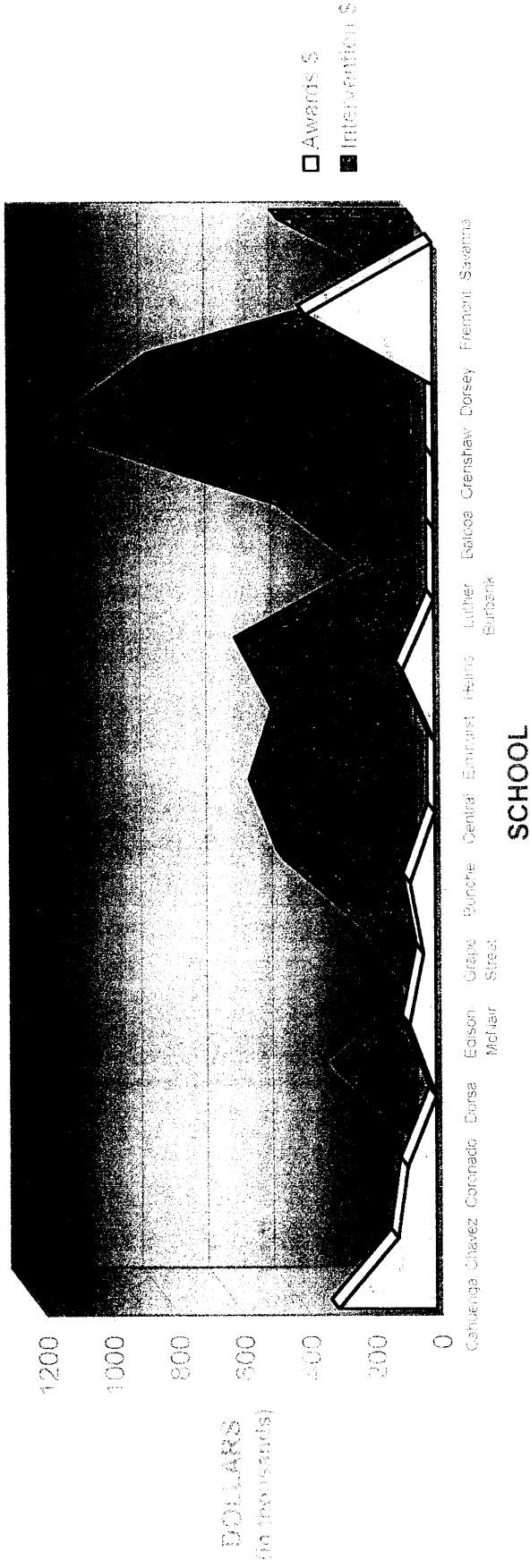


CHART 20
 Percents of Credentialed Teachers by Type for Schools Attended by Named Plaintiffs

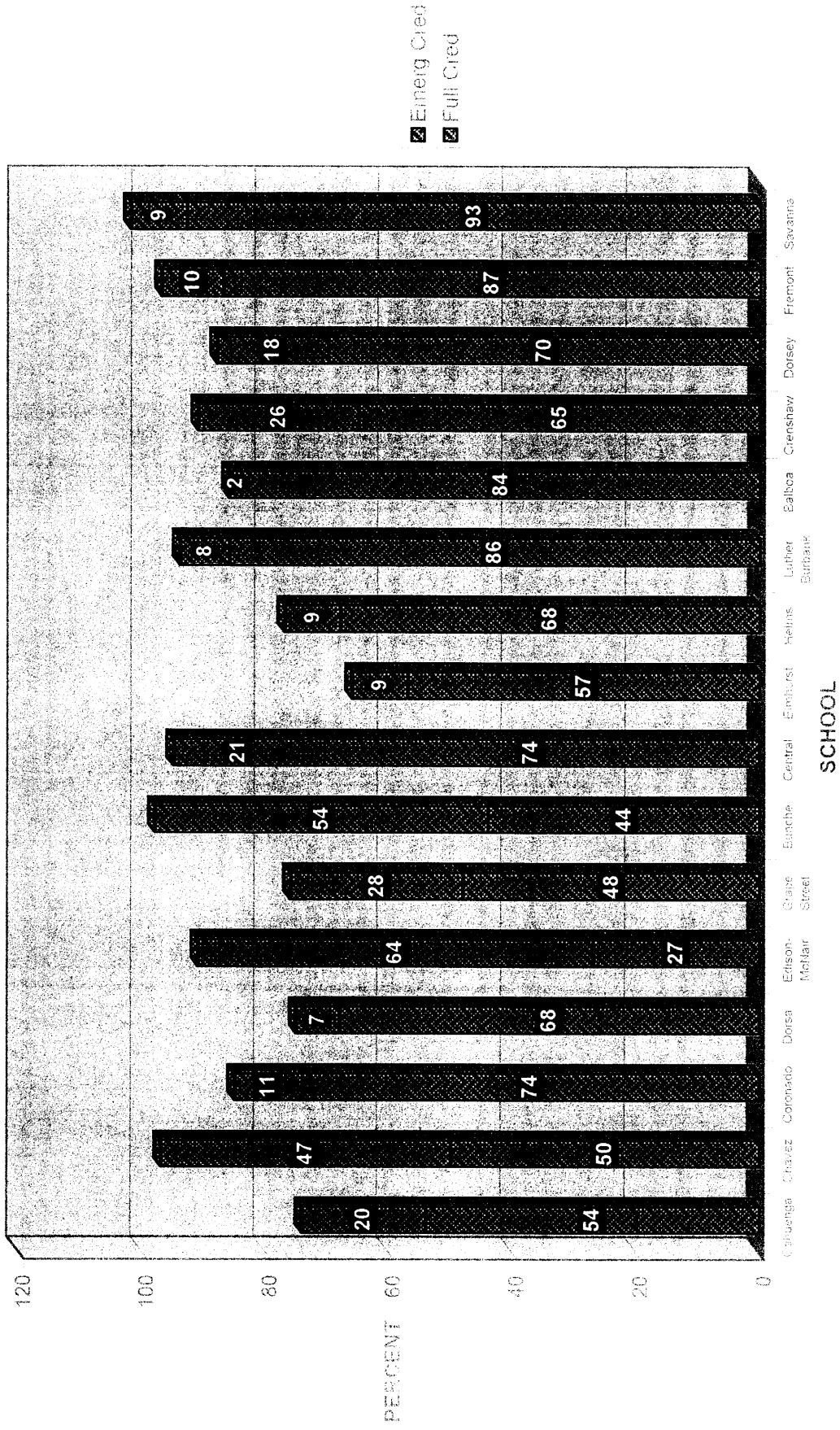


TABLE 1
Match of Stanford Test & California Test to California Content Standards*

GRADE	NO. OF STRANDS (STANDARDS)	STANFORD TEST			CALIFORNIA SURVEY TEST		
		PCT OF STRANDS ASSESSED	NO. OF TEST ITEMS	NO. OF AUGMENTED TEST ITEMS	PCT OF STANDARDS ASSESSED	NO. OF TEST ITEMS	PCT OF ALIGNED ITEMS
LANGUAGE ARTS							
2	16(46)	69%	192	40	28%	45	73%
3	17(49)	82%	162	40	37%	50	98%
4	17(52)	76%	162	55	29%	55	85%
5	19(42)	79%	162	55	38%	55	93%
6	19(50)	84%	162	55	30%	55	100%
7	18(45)	89%	162	55	29%	55	80%
8	17(48)	76%	162	55	35%	55	89%
9	15(63)	80%	132	55	18%	55	96%
10	15(63)	87%	132	55	58%	55	100%
11	14(56)	79%	132	55	16%	55	98%

GRADE	NO. OF STRANDS (STANDARDS)	STANFORD TEST			CALIFORNIA SURVEY TEST		
		PCT OF STRANDS ASSESSED	NO. OF TEST ITEMS	NO. OF AUGMENTED TEST ITEMS	PCT OF STANDARDS ASSESSED	NO. OF TEST ITEMS	PCT OF ALIGNED ITEMS
MATHEMATICS							
2	5(36)	100%	74	15	47%	26	100%
3	5(49)	100%	76	15	28%	30	87%
4	5(55)	100%	78	15	29%	30	90%
5	5(38)	100%	78	15	34%	30	70%
6	5(50)	100%	78	15	34%	30	87%
7	5(55)	100%	80	15	28%	30	96%
8 (Alg I)	8(55)	75%	82	15	11%	30	47%
9(Geom)	8(55)	75%	48	15	11%	25	36%
10 (Alg II)	8(55)	88%	48	15	11%	25	36%
11	8(55)	88%	48	15	11%	25	36%

* Source: Augmented Test Blueprint for STAR CST approved by SBE in 1999; Educational Testing Service, STAR Proposal, March 11, 2002, Component 1: Nationally Norm-Referenced Achievement Test, p. 16-17.

TABLE 2a
Statewide Stanford Test Summary Data

Stanford Test	READ (MATH) NPR FOR "AVG" STUDENT SCORE			READ (MATH) % ABOVE 50 TH NPR	
	All Students	ELLs >12 mo.	SES _D	African-American	Hispanic
Grade 2					
2002	52 (62)	35 (50)	39 (51)	45 (48)	38 (51)
2001	50 (59)	31 (44)	36 (46)	42 (44)	35 (46)
2000	48 (57)	29 (42)	34 (44)	40 (42)	31 (42)
1999	43 (50)	23 (35)*	29 (37)	33 (34)	25 (35)
1998	39 (43)	19 (27)*	NR (NR) [†]	NR (NR)	NR (NR)
Grade 4					
2002	50 (58)	24 (38)	34 (44)	36 (40)	30 (44)
2001	47 (54)	22 (33)	31 (40)	33 (36)	27 (39)
2000	45 (51)	20 (31)	29 (36)	30 (32)	24 (34)
1999	42 (44)	17 (25)*	25 (30)	27 (25)	21 (26)
1998	40 (39)	15 (21)*	NR (NR)	NR (NR)	NR (NR)
Grade 7					
2002	46 (54)	17 (30)	29 (39)	32 (31)	29 (34)
2001	46 (53)	16 (28)	29 (37)	32 (29)	27 (32)
2000	45 (51)	15 (26)	27 (34)	30 (26)	26 (29)
1999	43 (47)	14 (24)*	26 (32)	28 (24)	24 (25)
1998	41 (45)	12 (22)*	NR (NR)	NR (NR)	NR (NR)
Grade 10					
2002	33 (48)	9 (28)	18 (34)	19 (25)	16 (29)
2001	33 (47)	9 (27)	18 (33)	19 (23)	16 (27)
2000	33 (47)	9 (28)	18 (34)	18 (24)	15 (27)
1999	32 (45)	9 (27)*	18 (34)	17 (22)	15 (25)
1998	32 (43)	8 (25)*	NR (NR)	NR (NR)	NR (NR)

* All ELLs.

† NR = Not Reported.

TABLE 2b
Statewide Standards Test Summary Data

LANGUAGE ARTS (MATH)					
Standards Test	All Students	ELLs > 12 mo.	SES_D	African American	Hispanic
Grade 2	% PROFICIENT OR ABOVE				
2002	32 (43)	14 (27)	18 (30)	23 (29)	17 (30)
	PERCENT CORRECT[†]				
2001	59 (66)	48 (58)	51 (60)		
2000	58 (59)	46 (51)	49 (52)		
1999	55 (54)	43 (46)*	46 (48)		
Grade 4	% PROFICIENT OR ABOVE				
2002	36 (37)	10 (18)	19 (24)	24 (22)	19 (24)
	PERCENT CORRECT[†]				
2001	54 (60)	40 (49)	46 (53)		
2000	52 (54)	38 (44)	44 (47)		
1999	50 (47)	37 (38)*	42 (40)		
Grade 7	% PROFICIENT OR ABOVE				
2002	33 (30)	5 (8)	16 (16)	18 (13)	17 (15)
	PERCENT CORRECT[†]				
2001	58 (52)	43 (40)	50 (44)		
2000	58 (43)	42 (33)	49 (37)		
1999	57 (41)	41 (33)*	48 (35)		
Grade 10[†]	% PROFICIENT OR ABOVE				
2002	33 (22)	3 (9)	14 (12)	19 (9)	16 (10)
	PERCENT CORRECT[†]				
2001	57 (46)	39 (38)	47 (41)		
2000	56 (50)	39 (37)	47 (41)		
1999	56 (53)	40 (42)*	48 (46)		

* All ELLs.

† Math = Algebra I, Grades 8-11.

‡ NOT equated across years.

TABLE 2c
2002 Statewide Stanford Test Only versus Standards Tests Only APIs*

API Points	NPR/ Performance Level	SAT9 Reading	SAT9 Language	SAT9 Spelling	SAT9 Math	ELA Standards	Math Standards
80-99	Advanced		25.8%		30.1%		10.1%
60-79	Proficient		21.1%		19.9%		22.5%
40-59	Basic		16.9%		16.4%		27.7%
20-39	Below Basic		18.4%		18.6%		27.1%
1-19	Far Below Basic		17.8%		14.9%		12.6%
2002 State Base API							
Normal Weights					661.5		
SAT9 Only ('99 Weights)					682.8		
Standards Tests Only (60% ELA & 40% Math)					656.2		

* Statewide, grades 2-8 only, non-mobile students.

S A T 9	In API	Yes	Yes	Yes	Yes
	Performance Level	No	No	No	Yes
	Weight	0%	0%	0%	32%
S A T 9	In API	Yes	Yes	Yes	Yes
	Performance Level	No	No	Yes	Yes
	Weight	0%	0%	36%	48%

TABLE 3a
Data for Individual School Districts & Statewide Data

NPR FOR "AVERAGE" STUDENT															
Stanford Test	Inglewood USD			Oceanside USD			Hayward USD			Oakland USD			California		
Reading	ALL	ELL	SES_D	ALL	ELL	SES_D	ALL	ELL	SES_D	ALL	ELL	SES_D	ALL	ELL	SES_D
GRADE 4 2002	45	36	44	49	21	38	41	24	33	34	21	28	50	24	34
1998/99[†]	31	16	33	35	8	30	34	22	29	22	17	24	40	15	25
GROWTH	14	20	11	14	13	8	7	2	4	12	4	4	10	9	9
GRADE 10 2002	16	11	13	34	8	20	23	9	14	17	9	14	33	9	18
1998/99[†]	14	7	12*	26	2	19	24	12	17	18	9	18	32	8	18
GROWTH	2	4	1	8	6	1	-1	-3	-3	-1	0	-4	1	1	0
Mathematics	ALL	ELL	SES_D	ALL	ELL	SES_D	ALL	ELL	SES_D	ALL	ELL	SES_D	ALL	ELL	SES_D
GRADE 2 2002	73	74	74	65	50	58	57	50	52	47	41	42	62	50	51
1998/99[†]	50	43	63	41	18	40	37	34	33	31	31	35	43	27	37
GROWTH	23	31	11	24	32	18	20	16	19	16	10	7	19	23	14
GRADE 4 2002	56	53	56	61	38	53	50	38	42	38	31	33	58	38	44
1998/99[†]	37	24	45	35	11	35	29	23	28	22	24	26	39	21	30
GROWTH	19	29	11	26	27	18	21	15	14	16	7	7	19	17	14
GRADE 7 2002	32	26	32	55	26	43	46	32	39	36	30	32	54	30	39
1998/99[†]	26	20	24	37	11	29	37	26	32	29	24	27	45	23	32
GROWTH	6	6	8	18	15	14	9	6	7	7	6	5	9	8	7
GRADE 10 2002	26	22	25	50	31	41	38	28	33	34	30	34	48	28	34
1998/99[†]	22	20	23*	36	13	33	36	32	35	31	33	33	43	25	34
GROWTH	4	2	2	14	18	8	2	-4	-2	3	-3	1	5	3	0

[†] 1998 data for all students and ELLs; 1999 data for SES_D.

* 2000 data.

TABLE 3 b
Data for Individual Schools Attended by Named Plaintiffs

		NPR FOR "AVERAGE" STUDENT																					
Standard Test		Cahuenga Elem			Coronado Elem			Edison-McNair Acad			Bunche Middle			Luther Burbank Mid			Dorsey Sr High			Crenshaw Sr High			
	Reading	ALL	ELL	SES _b	ALL	ELL	SES _b	ALL	ELL	SES _b	ALL	ELL	SES _b	ALL	ELL	SES _b	ALL	ELL	SES _b	ALL	ELL	SES _b	
Grade 4	2002	48	46	48	34	31	34	16	11	17	25	33	25	32	15	34							
1998/99*		39	23	30	15	11	18	14	7	15*	13	31*	18*	26	11	26							
GROWTH		9	23	18	19	20	16	GROWTH	2	4	2	2	7	6	4	8							
Math																							
Grade 2	2002	82	82	82	57	50	57	34	28	32	34	43	34	39	25	38							
1998/99*		65	69	63	42	33	45	17	15	18	17	57	26	37	22	40							
GROWTH		17	13	19	45	47	42	GROWTH	17	11	14	-14	8	2	3	-2							
Grade 4	2002	70	69	82	55	57	55	25	21	25	34	43	34	39	25	38							
1998/99*		45	29	63	13	11	19	20	15	25	22	57	26	37	22	40							
GROWTH		25	40	19	42	46	36	GROWTH	5	6	0	12	8	2	3	-2							
Grade 10	2002																						
1998/99*																							
GROWTH																							

† 1998 data for all students and ELLs; 1999 data for SES_b.

* 2000 data.

TABLE 5
Comparison of California NAEP & Standards Tests Trend Data[†]

PERCENT PROFICIENT OR ABOVE						
	ALL STUDENTS	ASIAN	AFRICAN-AMERICAN	HISPANIC	WHITE	U.S.
READING						
GRADE 4						
1998	20	31	7	8	29	29
1994	18	29	8	6	25	28
1992	19	24	8	6	30	27
GRADE 8						
1998	22	27	12	9	36	31
MATH						
GRADE 4						
2000	15	25	2	5	25	25
1996	11	17	2	4	17	20
1992	12	21	2	4	19	17
GRADE 8						
2000	18	33	4	7	27	26
1996	17	29	2	5	28	23
1992	16	29	2	4	25	20
1990	12	20	3	3	29	15
ELA GRADE 4						
2002	36		24	19		---
2001	33		---	---		---
GRADE 8						
2002	32		17	15		---
2001	32		---	---		---
MATH GRADE 4						
2002	37		22	24		---
2001	---		---	---		---
GRADE 8*						
2002	39		15	18		---
2001	---		---	---		---

[†] Source: www.nces.ed.gov; www.cde.ca.gov

* Students taking Algebra I.

TABLE 6a
Summary of Statewide API School Performance[†]

PERCENTAGE OF SCHOOLS MEETING TARGETS				
	2001-2002*	2000-2001	1999-2000	
Elementary	60%	64%	79%	
Middle	39%	51%	60%	
High	29%	27%	41%	
ALL SCHOOLS	53%	57%	71%	
PERCENTAGE OF SCHOOLS WITH AN INCREASED SCHOOLWIDE API				
	2001-2002*	2000-2001	1999-2000	
Elementary	74%	78%	93%	
Middle	61%	72%	84%	
High	58%	53%	72%	
ALL SCHOOLS	69%	74%	89%	
PERCENTAGE OF SCHOOLS AT OR ABOVE API TARGET OF 800				
	2002*	2001	2000	1999
Elementary	24%	23%	20%	13%
Middle	17%	16%	14%	11%
High	6%	6%	6%	5%
ALL SCHOOLS	20%	20%	17%	12%
MEDIAN API SCORES BY SCHOOL TYPE				
	2001-2002*	2001*	2000	1999
	Growth	Base	Base	Base
Elementary	705	689	675	629
Middle	681	668	657	633
High	647	635	636	620
PERCENTAGE OF SCHOOLS ELIGIBLE FOR GOVERNOR'S PERFORMANCE AWARD				
	2001-2002*[‡]	2000-2001[‡]	1999-2000	
Elementary	46%	54%	75%	
Middle	26%	42%	56%	
High	21%	22%	38%	
ALL SCHOOLS	39%	48%	67%	

[†] Source: www.cde.ca.gov/api/. Adapted from API 2001-2002 Growth Results Media Packet.

* 2001-2002 was the first API cycle to include the California Standards Test for English Language Arts.

[‡] Beginning in 2001 the awards criteria changed affecting schools in the higher ranks. A minimum of 5 points growth is required schoolwide and 4 points growth for each numerically significant subgroup.

TABLE 6 b
API Awards Eligibility by Deciles & Reasons for NO 2002 School Growth API[▲]

PERCENT OF SCHOOLS ELIGIBLE FOR AWARDS			
DECILE RANK	2001-2002[†]	2000-2001*	1999-2000[‡]
1	57%	55%	62%
2	51%	58%	59%
3	45%	48%	63%
4	43%	47%	65%
5	36%	46%	66%
6	34%	49%	64%
7	32%**	47%**	68%
8	29%**	43%**	70%
9	33%**	47%**	73%
10	35%**	45%**	80%

REASONS WHY SOME SCHOOLS DID NOT RECEIVE 2002 API GROWTH RESULTS	Number of Schools
All Schools, Fall 2001	8,812
Schools Receiving 2002 Growth APIs	- 6,489
Schools Not Receiving 2002 Growth APIs	2,323
Newly Formed Schools (No Opportunity to Measure Growth)	46
Alternative Schools, Special Education Centers and Very Small Schools (fewer than 11 valid scores)	1,289
Schools in 2001 Base API Report Not Receiving 2002 Growth API	
➤ Data Corrections Pending from Test Publisher	762
➤ A Valid 2001 Base Score Does Not Exist due to adult testing irregularities in 2001	25
➤ Excessive Parent Waivers in 2001 or 2002	97
➤ Not a significant Percentage of 2001 STAR Scores in a Content Area	7
➤ Not a Significant Percentage of 2002 STAR Scores in a Content Area	21
➤ Unresolved Data Discrepancies	1
➤ Testing Irregularities Reported by Districts in 2002	28
➤ API Not Comparable (Reported by District)	23
➤ No 2002 Test Results	23
➤ Missing Some STAR Test Results in 2002	1
Subtotal	988
TOTAL	2,323

[▲] Source: www.cde.ca.gov/api/. Adapted from API 2001-2002 Growth Results Media Packet.

[†] January 2001 API Base deciles.

* January 2000 API Base deciles.

[‡] January 1999 API Base deciles.

** Beginning in 2001 the awards criteria changed affecting schools in the higher ranks. A minimum of 5 points growth is required schoolwide and 4 points growth for each numerically significant subgroup.

TABLE 7
Ratings of Selected State Standards & Accountability Systems¹

2000 Fordham Foundation National Report Card State Standards [†]									2003 Ed Week	
STATE	Eng LA/Read	History	Geography	Math	Science	Overall		Accountability	Standards & Acct [†]	
CA*	A	A	C	A	A	3.6	A-	Strong	88	B+
AZ	B	A	B	B	A	3.4	B+	Weak	77	C+
SC*	B	C	A	B	B	3.0	B	Strong	86	B
TX*	B	B	A	B	C	3.0	B	Strong	79	C+
AL*	A	B	B	B	D	2.8	B-	Strong	79	C+
NC*	B	D	C	A	A	2.8	B-	Strong	81	B-
MA	A	B	D	D	A	2.6	B-	Weak	92	A-
SD	C	C	C	A	B	2.6	B-	Weak	75	C
≈	≈	≈	≈	≈	≈	≈	≈	≈	≈	≈
RI	F	—	—	F	A	1.3	D+	Weak	68	D+
US	C-	D+	C-	C	C	1.7	C-	Weak	78	C+

¹ States listed in order by Fordham standards rank. All 8 states with honors (A or B) grades plus RI (accountability system cited as good model by the Russell Report at 47) are included.

* Honor Roll: Solid Standards; Strong Accountability ("Only these five states can claim to be doing standards-based reform well."). Finn, C. & Petrilli, M. (Ed.), *The State of State Standards 2000*, The Thomas B. Fordham Foundation, Jan. 2000, p. 3.

† Fordham Standards Criteria: English Language Arts/Reading – 34 criteria in 5 categories of purpose and expectations, organization, disciplinary coverage, quality, absence of anti-literary expectations; History – 15 criteria in 5 categories of clarity, organization, historical soundness, content in US, European and world history, absence of manipulation; Geography – 6 general criteria plus comprehensiveness and rigor of content and skills; Mathematics – 9 indicators in 4 groups of clarity, content, math reasoning, absence of negative qualities; Science – 25 criteria in 5 categories of purpose and expectations, organization, coverage and content, quality, absence of negatives. *Id.* at xi, 1, 129, 137, 143, 151, 155.

† Fordham Accountability Criteria: report cards that include test scores, rewards for successful schools, school ratings of academic performance, sanctions for failing schools that are utilized. *Id.* at 3, 175.

† Ed Week Criteria: 15% adoption of standards in English, math, science and social studies; 25% standards clear, specific and grounded in content; 28% aligned criterion-referenced assessments in four core subjects using a variety of item types (multiple-choice, short-answer, extended-response English, extended-response other subject and portfolio) at each level (elementary, middle, high school); 2% participation in NAEP; 30% accountability components – report cards, ratings, assistance, rewards and sanctions – in place. *Quality Counts 2003*, Ed Wk, XXII (17), Jan. 9, 2003, p. 84-85, 102.

AFT RATINGS*	STANDARDS SCORE	STANDARDS RANK	ASSESSMENTS SCORE	ASSESSMENTS RANK
California	24	1	16 [†]	20
Arizona	24	1	12	32
South Carolina	23	6	12	32
Texas	19	33	20	12
Alabama	21	24	14	28
North Carolina	18	37	10	43
Massachusetts	23	6	23	10
South Dakota	22	17	12	32
Rhode Island	12	49	6	47

* American Federation of Teachers. *Making Standards Matter*. 2001. www.aft.org. Maximum 24 points for Elementary, Middle & High School Standards in English, Math, Science & Social Studies using AFT rating symbols with 2 points for • 1 point for • Standards = clear, specific & grounded in content, assessments = aligned to standards

† Science & Social Studies judged *not* aligned at Elementary & Middle School because Standards Tests not yet implemented.

TABLE 8
Fordham Evaluation of State Standards & Accountability Systems*

	SOLID STANDARDS (A OR B AVERAGE)	MEDIOCRE STANDARDS (C AVERAGE)	INFERIOR OR NO STANDARDS (D OR F AVERAGE OR (NC))
Strong Accountability	<u><i>The Honor Roll</i></u> Alabama California North Carolina South Carolina Texas	<u><i>Shaky Foundations</i></u> Florida Illinois Indiana Kansas Maryland Nevada New York Oklahoma Virginia West Virginia	<u><i>Trouble Ahead</i></u> Kentucky New Mexico
Weak Accountability	<u><i>Unrealized Potential</i></u> Arizona Massachusetts South Dakota	<u><i>Going Through the Motions</i></u> Delaware Georgia Louisiana Mississippi Nebraska New Hampshire Ohio Utah Wisconsin	<u><i>Irresponsible States</i></u> Alaska Arkansas Colorado Connecticut Hawaii Idaho Iowa Maine Michigan Minnesota Missouri Montana New Jersey North Dakota Oregon Pennsylvania Rhode Island Tennessee Vermont Washington Wyoming

* Source: Finn C. & Petrilli, M. (Ed.), *The State of State Standards 2000*, The Thomas B. Fordham Foundation, January 2000, p. 3, Figure 1, emphasis added.

The Honor Roll. Only these 5 states can claim to be doing standards-based reform well.

Unrealized Potential. These 3 states have great academic standards. Too bad they don't count for much.

Shaky Foundations. These 10 states have built school-based accountability on a relatively weak foundation.

Going Through the Motions. With mediocre standards and little or no accountability, these 9 states give lip-service to standards-based reform, but not much else.

Trouble Ahead. With high stakes attached to bad standards, Kentucky and New Mexico might inadvertently destroy some great schools – and push all schools towards more nonsense.

Irresponsible States. These 21 states cannot claim to embrace standards-based reform.

TABLE 9
Demographic Data for Selected States*

	# Public Schools	# Pub Sch Teachers	# Pre-K-12 Students	Minority	Poverty	Disabled	ELL	Annual Expenditures (in billions)	Per Pupil Spending (adjusted)
CA	8,757	305,000	6,248,000	63%	23%	11%	25%	\$ 43.0	\$6,161
AZ	1,633	46,000	904,000	47%	23%	10%	15%	\$ 4.9	\$ 5,487
SC	1,067	46,000	648,000	45%	22%	15%	1%	\$ 4.7	\$ 7,930
TX	7,519	281,000	4,128,000	58%	22%	12%	14%	\$ 28.2	\$ 7,248
AL	1,380	47,000	726,000	39%	23%	14%	1%	\$ 4.3	\$ 6,652
NC	2,192	84,000	1,304,000	39%	19%	14%	3%	\$ 8.6	\$ 7,170
MA	1,898	69,000	980,000	24%	14%	16%	5%	\$ 9.7	\$ 8,429
SD	756	9,100	127,000	14%	18%	13%	4%	\$ 0.8	\$ 7,540
RI	320	10,000	158,000	26%	16%	19%	7%	\$ 1.6	\$9,265
US	90,640	2,988,000	47,576,000	39%	19%	13%	8%	\$358.0	\$7,524

* Same states as Table 1. Source: *Quality Counts 2003*, Education Week, XXII (17), January 9, 2003, p. 98, 106, 107, 111, 114, 117, 120, 138, 156, 164, 165, 166, 169.

TABLE 10
Test Score Data for Selected States*

State [†]	NAEP Percent Proficient or Above 2000 1998				2002 State CRT Percent Proficient of Above [‡]				1999 TIMSS	2001 SAT			2002 ACT	
	4 th Math	8 th Math	4 th Read	8 th Read	4 th Math	8 th Math	4 th ELA	8 th ELA	8 th Math	% Tested	V	Q	% Tested	TOTAL
CA	15	18	20	22	37	29 [‡]	36	32	—	51	498	517	13	21.4
AA	2	4	7	12	22	11 [‡]	24	17						
H	5	7	8	9	24	12 [‡]	19	15						
AZ	17	21	22	28	53 [‡]	20	71 [‡]	56	—	34	523	525	26	21.3
AA	5	8	10	11	—	—	—	—						
H	6	8	8	13	—	—	—	—						
SC	18	18	22	22	26 [‡]	18 [‡]	37 [‡]	24 [‡]	502	57	486	488	32	19.2
AA	4	4	10	9	13 [‡]	13 [‡]	15 [‡]	15 [‡]						
H	12	9	10	9	24 [‡]	24 [‡]	25 [‡]	25 [‡]						
TX	27	24	29	28	94 [‡]	93 [‡]	93 [‡]	94 [‡]	516	53	493	499	32	20.1
AA	12	6	11	12	89 [‡]	87 [‡]	87 [‡]	92 [‡]						
H	14	14	15	15	93 [‡]	90 [‡]	90 [‡]	91 [‡]						
AL	14	16	24	21	70 [‡]	72 [‡]	64 [‡]	64 [‡]	—	9	559	554	71	20.1
AA	4	4	8	7	—	—	—	—						
H	5	6	7	12	—	—	—	—						
NC	28	30	28	31	89	82	77	85	495	65	493	499	13	19.9
AA	9	7	11	13	79	68	62	72						
H	13	18	13	12	86	71	67	71						
MA	33	32	37	36	39	34	54	—	513	79	511	515	8	21.9
AA	7	8	12	15	14	11	26	—						
H	10	14	14	15	12	8	23	—						
SD	—	—	—	—	65 [‡]	70 [‡]	65 [‡]	65 [‡]	—	4	577	582	71	21.4
AA	—	—	—	—	30 [‡]	30 [‡]	40 [‡]	40 [‡]						
H	—	—	—	—	27 [‡]	27 [‡]	39 [‡]	39 [‡]						
RI	23	24	32	30	37 [‡]	27 [‡]	64 [‡]	37 [‡]	—	71	501	499	6	21.9
AA	4	6	12	16	16 [‡]	10 [‡]	50 [‡]	19 [‡]						
H	5	4	8	10	15 [‡]	9 [‡]	45 [‡]	17 [‡]						
US	25	26	29	31					502	45	506	514	39	20.8
AA	5	5	9	11										
H	10	9	12	14										

* Same states as Table 7. Source: NAEP – *Quality Counts 2003*, Education Week, XXII (17), January 9, 2003, p. 80, 82 & www.nces.us.gov; State Tests – state websites & www.ccsso.org for 1999-2000; TIMSS – www.ustimss.msu.edu; SAT & ACT – Nation's Report Card, www.nces.us.gov.
[†] AA= African-American; H=Hispanic.
[‡] TX: New, more challenging tests administered in spring 2003; percents will decrease significantly.
[§] CA: Average across subject tests (e.g., Algebra I, Geometry, Algebra II, etc.)
[¶] AZ: 4th = elementary average from 2000 CCSSO data; [‡]SC: 2001 data, AA & H average all grades tested; [§]AL: 2000 data; [¶]SD: 2000 CCSSO data, elementary & middle school average, AA & H average all grades tested; [¶]RI: 2001 data, math = average skills, concepts & problem solving, ELA = average basic understanding & analysis and interpretation, AA & H estimated from 1999-2001 summary elementary & middle school data.

TABLE 11
California Public Opinion Poll Data*

Statement/Question	Response		
	Agree	Disagree	No answer
There is just no way to hold schools accountable for student progress.	22%	76%	3%
We are not doing a good enough job holding schools accountable for the quality of education they provide.	70%	26%	5%
Expectations of student performances are far too low in California.	71%	23%	6%
Public school students have to take too many tests.	29%	60%	12%
Do you favor or oppose having standardized tests as part of California's public education system?	Favor 78%	Oppose 20%	No answer 3%
Do you favor or oppose requiring students to pass the California High School Exit Exam in order to receive a high school diploma from a public school?	77%	20%	4%
After you read each statement, please tell me whether it makes you more inclined or less inclined to favor standardized testing.	More inclined	Less inclined	No answer
Standardized tests hold schools accountable for teaching students what they need to know.	78%	20%	3%
Public reporting of standardized testing results puts pressure on schools to do better.	80%	17%	4%
After you read each statement, please tell me whether it makes you more inclined or less inclined to oppose standardized testing.	More inclined	Less inclined	No answer
Public schools should be given more money before they are held accountable for test results.	37%	53%	10%
Which is more important to know: if students are learning state standards or how students are doing compared to other students across the country?	Learning state standards 45%	National comparison 36%	Both (No answer) 11% (8%)
How effective do you think the annual STAR tests are in determining whether California's public schools are improving?	Effective 62%	Ineffective 22%	No answer 16%
Would you say that having the same learning standards for all schools statewide is a good idea or not?	Good idea 86%	Not a good idea 9%	No answer 5%
Should we ... ?	Keep momentum going with more reforms 58%	Take a break so recent reforms can be fully implemented 28%	No answer 14%

* Random sample telephone survey of 800 frequent California voters conducted in August 2002 by Fairbank, Maslin, Maullin & Associates for the California Business for Education Excellence Foundation. The margin of error is $\pm 3.5\%$.

TABLE 12
2001 Mean Number of Teachers with Full & Emergency Credentials
and SES_D API Gains for Elementary Schools (Grades 2-6) by Decile[†]

2000-01 DECILE	MEAN SCHOOL PERCENT		SES _D API GROWTH		SES _D API GROWTH IN HIGH SES _D SCHOOLS*	
	FULL CRED	ECRED	ECRED > 15	ECRED = 0	ECRED > 15	ECRED = 0
1	72.6	20.3	41.1	24.1	41.1	24.1
2	79.8	16.0	32.8	34.2	32.5	34.2
3	82.8	13.7	28.3	29.1	28.6	29.4
4	85.6	11.7	28.5	32.4	28.5	32.4
5	88.0	9.9	27.7	24.4	28.4	19.4
6	91.1	7.5	15.6	19.7	16.0	14.6
7	92.5	6.5	21.1	20.5	18.4	6.3
8	94.1	5.4	13.7	12.5	2.8	11.6
9	94.4	5.0	-14.0	11.7	-17.9	4.0
10	95.5	4.3	8.5	5.9	—	—

[†] Source: Rogosa, D., *Teacher Credentials and Student Progress: What do the data say?*, Stanford University, December 2002, Tables 2B, 2C, & 4C.

* Schools with SES_D > 50%.

TABLE 13 a
Summary API Data for CA Elementary Schools Attended by Named Plaintiffs*

API Growth Report						Base Yr Rank		Met Growth Target			Funding†		
YEAR	% TESTED	N	API	TARGET	GROWTH	STATE	SIM SCHLS	SCHL	COMP	IMP	GROUPS†	AWARDS	INTERVENTION
Cahuenga Elementary (Los Angeles Unified District)						Los Angeles County							
2002	99	779	727	5	30	6	10	Y	Y		A, H, S _D	Eligible	
2001	99	624	694	9	67	4	10	Y	Y			\$ 61,584	
2000	100	584	627	10	27	5	10	Y	Y			\$ 237,901	
'99 Base			(600)										
Chavez Academy (Ravenswood City Elem District)						San Mateo County							
2002	Not Reported – Parents excused ≥ 10%										No		
2001	Not Reported – Parents excused ≥ 20%										No		
2000	98	425	488	20	80	1	2	Y	Y		H, S _D	\$ 116,653	
'99 Base			(408)										
Coronado Elementary (W. Contra Costa Unified District)						Contra Costa County							
2002	100	244	591	13	50	2	6	Y	Y		AA, H, S _D	Eligible	
2001	100	242	543	18	109	1	1	Y	Y			\$ 20,870	
2000	95	241	434	21	53	1	1	Y	Y			\$ 66,278	
'99 Base			(381)										
Dorsa Elementary (Alum Rock Union Elem District)						Santa Clara County							
2002	100	313	524	16	42	1	1	Y	Y		H, S _D	Eligible	\$238,800
2001	Not Reported – Validity uncertain because demographic data not corrected as promised										No	NEW	
2000	98	302	433	19	16	1	1	N	Y		H, S _D	No	
'99 Base			(417)										
Edison-McNair Academy (Ravenswood City Elem District)						San Mateo County							
2002	99	496	491	18	41	1	1	Y	Y		AA, H, S _D	Eligible	
2001	98	453	436	18	0	1	4	N	N			No	
2000	96	426	436	21	52	1	2	Y	Y			\$ 80,401	
'99 Base			(384)										
Grape Street Elementary (Los Angeles Unified District)						Los Angeles County							
2002	100	421	555	13	20	2	7	Y	N		AA, H, S _D	No	\$151,800
2001	99	395	526	20	117	1	1	Y	Y			\$ 36,365	CONT
2000	100	421	409	22	54	1	1	Y	N			No	
'99 Base			(355)										

* Source: www.cde.ca.gov/api/; www.cde.ca.gov/ope/; www.cde.ca.gov/i/iusp/.

† AA=African-American, A=Asian, F=Filipino, H=Hispanic, S_D=Socioeconomically Disadvantaged.

‡ Awards = (GPA + AB1114 + SB1557) in 2000 (see Chart 10); GPA in 2001; Nonmonetary in 2002; Intervention = 2002-2003 funding (continuation or new).

TABLE 13 b
Summary API Data for CA Middle Schools Attended by Named Plaintiffs*

API Growth Report						Base Yr Rank		Met Growth Target			Funding [†]		
YEAR	% TESTED	N	API	TARGET	GROWTH	STATE	SIM SCHLS	SCHL	COMP	IMP	GROUPS [‡]	AWARDS	INTERVENTION
Bunche Middle (Contra Costa Unified District)						Alameda County							
2002	100	952	532	15	25	1	7	Y	Y		AA, H, Sp	Eligible	\$401,600
2001	100	708	500	19	80	1	3	Y	Y			\$ 75,892	NEW
2000	100	850	420	19	1	1	4	N	N			No	
99 Base			(419)										
Central Junior High (Piedmont Unified District)						Contra Costa County							
2002	97	966	562	15	52	1	2	Y	Y		AA, H, W, Sp	Eligible	\$485,600
2001	100	1068	490	16	10	1	1	N	N			No	NEW
2000	99	867	480	15	-17	2	4	N	N			No	
99 Base			(497)										
Elmhurst Middle (Oakland Unified District)						Alameda County							
2002	92	902	459	18	14	1	1	N	N		AA, H, Sp	No	\$415,200
2001	91	869	425	19	11	1	2	N	N			No	NEW
2000	87	779	446	18	32	1	5	N	N			No	
99 Base			(414)										
Helms Middle (W. Contra Costa Unified District)						Contra Costa County							
2002	100	1216	474	16	7	1	3	N	N		AA, A, H, Sp	No	\$527,200
2001	97	1012	466	19	36	1	2	Y	Y			\$ 96,130	NEW
2000	98	1034	430	20	20	1	2	Y	N				
99 Base			(410)										
Luther Burbank Middle (San Francisco Unified District)						San Francisco County							
2002	98	571	572	13	25	2	1	Y	N		AA, A, H, Sp	No	\$137,400
2001	97	588	543	13	7	2	1	N	N		A, H, Sp	No	CONT
2000	95	600	536	13	-7	3	1	N	N		AA, A, H, Sp	No	
99 Base			(543)										

* Source: www.cde.ca.gov/api/; www.cde.ca.gov/ope/; www.cde.ca.gov/i/usp/.

† AA=African-American, A=Asian, F=Filipino, H=Hispanic, Sp=Socioeconomically Disadvantaged.

‡ Awards = (GPA + AB1114 + SB1557) in 2000 (see Chart 10); GPA in 2001; Nonmonetary in 2002; Intervention = 2002-2003 funding (continuation or new).

TABLE 13c
Summary API Data for CA High Schools Attended by Named Plaintiffs*

API Growth Report						Base Yr Rank		Met Growth Target			Funding [‡]		
YEAR	% TESTED	N	API	TARGET	GROWTH	STATE	SIM SCHLS	SCHL	COMP	IMP	GROUPS	AWARDS	INTERVENTION
Balboa High						(San Francisco Unified District)		San Francisco County					
2002	97	632	449	18	9	1	1	N	N		AA,F,H,S ₀	No	\$ 418,400
2001	89	555	435	18	-10	1	1	N	N			No	NEW
2000			(445)									—	
Not open in 1998-99 school year													
Crenshaw Senior High						(Los Angeles Unified District)		Los Angeles County					
2002	84	1668	463	17	8	1	3	N	N		AA,H,S ₀	No	\$1,089,600
2001	91	1671	445	17	7	1	3	N	N			No	NEW
2000	93	1687	452	17	-7	1	5	N	N			No	
'99 Base (459)													
Dorsey Senior High						(Los Angeles Unified District)		Los Angeles County					
2002	88	1255	460	18	18	1	3	Y	Y		AA,H,S ₀	No	\$ 789,200
2001	82	1152	431	18	-9	1	3	N	N			No	NEW
2000	92	1144	440	18	1	1	3	N	N			No	
'99 Base (439)													
Fremont High						(Fremont Union High District)		Santa Clara County					
2002	94	1070	674	6	2	7	9	N	N		A,F,H,W,S ₀	No	—
2001	100	1060	677	8	27	6	7	Y	Y			\$ 97,632	—
2000	93	655	643	9	16	6	6	Y	Y			\$ 288,495	—
'99 Base (627)													
Savanna High						(Anaheim Union High District)		Orange County					
2002	100	1442	586	11	0	4	3	N	N		A,H,W,S ₀	No	\$412,400
2001	100	1343	585	10	-14	4	5	N	N			No	CONT
2000	100	1298	599	11	26	4	4	Y	N			No	
'99 Base (573)													

* Source: www.cde.ca.gov/api/; www.cde.ca.gov/ope/; www.cde.ca.gov/i/usp/

¹ AA=African-American, A=Asian, F=Filipino, H=Hispanic, S₀=Socioeconomically Disadvantaged.

[‡] Awards = (GPA + AB1114 + SB1557) in 2000 (see Chart 10); GPA in 2001; Nonmonetary in 2002; Intervention = 2002-2003 funding (continuation or new).

TABLE 14
Summary Demographic Data for CA Schools Attended by Named Plaintiffs

	Ethnic Percent					# Parent Excluded	School	Mobility†		Percent Teachers	
	Af-Am	Asian	Filip	Hisp	SESp			ELL	School	District	Full Cred
ELEMENTARY											
Cahuenga	2	25	5	67	99	79	12	10%	4%	54	20
Chavez	10	3	0	74	100	71	75	18%	---	50	47
Coronado	52	0	1	44	99	39	0	20%	9%	74	11
Dorsa	1	4	2	87	95	65	0	16%	13%	68	7
Edison-McNair	24	0	1	64	77	67	0	14%	8%	27	64
Grape Street	33	0	0	66	100	48	0	18%	6%	48	28
MIDDLE/JR HIGH											
Bunche	18	0	0	77	100	55	0	8%	12%	44	54
Central	33	3	8	39	60	26	2	13%	12%	74	21
Elmhurst	49	4	0	44	80	41	26	12%	6%	57	9
Helms	23	10	3	61	91	53	0	38%	9%	68	9
Luther Burbank	20	18	14	35	70	23	4	18%	5%	86	8
HIGH SCHOOL											
Balboa	23	14	21	27	52	29	2	26%	11%	84	2
Grenshaw	74	0	0	25	71	9	56	18%	8%	65	26
Dorsey	57	0	0	40	71	20	15	19%	7%	70	18
Fremont	4	16	15	34	22	27	10	10%	38%	87	10
Savanna	5	12	4	49	46	21	7	7%	7%	93	9

‡ Percent of students first attending in current year.

* Tested students who attended this school the previous year.

† 1999 data.

TABLE 15
Award Probabilities & False Positives*

Probability of Award for Elementary Schools

SMALL SCHOOL ADVANTAGE?	TRUE API IMPROVEMENT		
	0	29	41
0	.13	.45	.60
1	.09	.62	.80
2	.10	.71	.89
3	.08	.82	.96

False Positive Results by School Type

each cell: average probability *no* improvement for award schools
(expected number of no improvement schools w/awards)

AWARD CYCLE	ELEMENTARY	MIDDLE	HIGH
1996-2000	.01 (35)	.02 (13)	.03 (9)
2000-2004	.03 (75)	.03 (14)	.04 (8)

* Source: Rogosa, D., *Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger "Volatility in School Test Scores"*, Stanford University, October 2002, Section 4, Parts A & C.