

Deceived and Confused:
An Attempt to Reconcile the Numbers in the Public Forum on School
Accountability Report, "A Better Student Data System for California"

David Rogosa and Edward Haertel
Stanford University
July 2003

Acknowledgements

Thanks to Ghassan Ghandour for the usual computing wizardry and to the students in Stanford University Education 351, Spring 2003 for helpful discussion.

For comments or inquiries contact D. Rogosa, rag at stat.stanford.edu

Deceived and Confused: An Attempt to Reconcile the Numbers in the Public Forum on School Accountability Report, "A Better Student Data System for California"

Summary

The Public Forum on School Accountability (PFSA) Report uses elementary school student test data from Los Angeles Unified (LAUSD) to demonstrate the advantages of longitudinally linked test scores (i.e., student identifiers that can be tracked over multiple years of testing). Closer examination reveals that the data presentations in the PFSA report *A Better Student Data System for California* are exceptionally misleading. Because acceptance of the PFSA claims would lead to an extreme lack in confidence in the results of the past 4 years of California API data, it is important to counter the PFSA presentation.

1. The Public Forum on School Accountability Report, "A Better Student Data System for California"

Background.

This report *A Better Student Data System for California*, April 2003 appears with impressive credentials. The report identifies the principal author as Philip Kaufman of MPR Associates. Members of the *Data Systems Panel* are listed as: Phil Daro, Don Barfield, Kathleen Barfield, George Bohrnstedt, Russ Brawn, Camille Esch, Neal Finkelstein, Ron Fox, Bob Friedman, Mike Garet, Pete Goldschmidt, Laura Hamilton, Phillip Kaufman, Don McLaughlin, Bill Padia, Russ Rumberger, Patrick Shields, Brian Stecher, Brad Strong. In addition, Foundation support was identified from: The William and Flora Hewlett Foundation, The James Irvine Foundation, The Stuart Foundation, The San Francisco Foundation, The East Bay Community Foundation, The S.H. Cowell Foundation, The Clarence E. Heller Charitable Foundation, The Walter S. Johnson Foundation, The Wallace Gerbode Foundation.

PFSA Results.

The key data analysis assertions from the PFSA report are contained on pages 10-11:

Examining STAR reading scores from 2000 to 2001 we measured the percentage of times cross-sectional data and longitudinal analysis yielded contradictory information about a school's progress. Figure 4 displays the results. [The figure 4 pie chart has the header]

Figure 4: PERCENTAGE AGREEMENT ON SCHOOL PROGRESS BETWEEN THE CALIFORNIA STATE API AND LONGITUDINAL DATA FROM THE LAUSD SCHOOLS

These data show that the API, based on solely quasi-longitudinal analyses, yields a reverse conclusion about school progress about 40 percent of the time compared to true longitudinal analysis. Of the 406 schools in LAUSD, about 8 percent showed increases in their achievement when examined longitudinally, while the State API had their scores declining. About 30 percent had longitudinal declines and API increases.

These conclusions have been repeated in various policy forums such as the April 2003 EdSource 26th Annual Forum, *California School Reforms: Show Me the Data!* (EdSource, 2003, p.2)

Taken at face value these assertions are devastating for the current uses, both descriptive and accountability (i.e., awards and sanctions), of the California API. Does the currently constituted API provide an incorrect indication of whether or not a school is showing year-to-year improvement for four out of ten California schools? Also, these issues are important for the implementation of California Senate Bill 1453 (Alpert), signed by Gov. Gray Davis in September 2002.

To bolster the criticisms of the API, the PFSA report (p.9) cites and quotes from Linn and Haug (2002), who claim that year-to-year changes in school scores are "quite unstable." However, Rogosa (2003) demonstrates that the methodology used by Linn and Haug has no utility, and that the empirical conclusions of volatility or stability based on those methods have no credibility. Furthermore, in Rogosa (2003, section 2) API scores for California elementary schools are seen to display good consistency in year-to-year improvement over the 1999-2002 period. And the consistency in year-to-year improvement is even stronger for the Socioeconomically Disadvantaged (SD) API subgroup scores.

The best way to illustrate the credibility of the PFSA assertions is to examine the small bit of actual numerical information that is provided in their report. Figure 3 of the PFSA report (reconstructed below) provides the exemplar school examples for the discordance between the current practice (which lacks longitudinal tracking of students) and the results obtained from the LAUSD longitudinally matched samples.

PFSA Figure 3:
 READING SCORE GAINS AT SEVERAL LAUSD SCHOOLS, 2000-2001

SCHOOL	CROSS-SECTIONAL GAIN*	LONGITUDINAL GAIN**
CLIFFORD ELEMENTARY	-15	0.6
CLOVER ELEMENTARY	-10	2.2
KITTRIDGE ELEMENTARY	4	-2.2
LANAI ELEMENTARY	27	-0.3

* Gain in percentage of students at or above the 50th percentile.

** NCE gain.

The main section of this report examines the test scores for each of these four PFSA exemplar schools. To set the record straight requires a lot of detail in the examples. In addition, for the full collection of LAUSD elementary schools an overall comparison of CDE API with a school summary based on the matched longitudinal sample demonstrates that the 40% discordance in school progress claimed by PFSA is more like 8%.

2. Comparing Cross-sectional and Longitudinal Results: The Four School Examples.

An interesting and clear "apples-to-apples" comparison would be between the API as currently constituted and a school-level measure based on the longitudinally matched sample. Despite the assertions in the PFSA report that the comparison being made was between the California API and a longitudinal gain measure, the one thing that is clear is that no tabulation of the California API was done. The examples below uncover what was done and why no credibility should be given to the PFSA assertions regarding the currently constituted API. For the four school examples in PFSA Figure 3 the first two indicate large decline from "cross-sectional gain" and increases from "longitudinal gain". The third and fourth example schools are the opposite: cross-sectional improvement, longitudinal decline.

A. Clifford Street Elementary, 19647336016513

Clifford is identified as having extreme decline between 2000 and 2001, yet examination of this school's API report shows year 2000 API of 592 and year 2001 growth API of 705, an increase of 113 API points! Only 13 schools in California made API gains that large between 2000 and 2001 (and, interestingly, 5 of these 13 were in LAUSD).

So where does the PFSA -15 number (15 point decline in the PAC metric) for Clifford come from? The answer is obtained from examining the STAR reading reports for Clifford.

Grade	SAT-9 Reading, STAR Report					
	n	Year 2000		n	Year 2001	
		PAC50	PAC25		PAC50	PAC25
2	24	17	54	29	48	86
3	31	32	58	24	17	54
4	23	35	70	28	43	79
5	24	21	63	19	32	63

First off, school level summaries based on the STAR Reading data again show strong year-to-year improvement, not the decline claimed by PFSA. The school-wide year 2000 PAC50 for Reading is 26.5 and the year 2001 PAC50 for Reading is 36, a strong 10 point year-to-year improvement.

Another, school-level summary (and one seemingly more appropriate for Clifford) is PAC25; school-wide year 2000 PAC25 for Reading is 60.8 and the year 2001 PAC25 for Reading is 72, again very strong year-to-year improvement.

What is apparent from the STAR Reading display is that the grade 2 cohort in year 2000 has poorer test performance than the other grades, and the 24 students in grade 2 in 2000 have identical PAC results to the (not necessarily the same) 24 students in grade 3 year 2001. As the grade-level cohorts move from year 2000 to 2001 some improve (grade 3, yr 2000, grade 4 year 2001) and others decline (grade 4, yr 2000, grade 5 year 2001). But the PFSA tabulation does not track an artificial cohort; instead PFSA takes a single grade in 2000 and compares with a single grade in 2001. That cross-sectional comparison would show improvement for Clifford in grades 2,4, or 5. But comparing the grade 3 year 2000 PAC50 versus grade 3 year 2001 PAC50 produces $17 - 32 = -15$, and thus Clifford is labeled a cross-sectional decliner.

For many good reasons, California does not report API scores for individual grades. Therefore it is strange, at the very least, for PFSA to assert data analysis conclusions about the API (see the claims for their Figure 4) based on the (very flawed) single-grade data displays. Moreover PAC50 is **not** the API, which makes even stranger the claimed equivalence of a single-grade PAC50 on a single subject to what PFSA repeatedly cites as the "California State API." Bizarre at the very least, and undeniably deceptive in consequence.

To get closer to what is done in the API, the putative interest of the PFSA comparisons, consider results for the subsets of students included in the 2000 and 2001 API. (Note: students included in STAR are excluded from the API, most often, for district or school mobility.) On every school-wide measure, Clifford, contrary to the PFSA classification, shows strong improvement from year 2000 to 2001. Instead of the 15 point decline in PAC50 claimed by PFSA (Fig. 3), here PAC50 actually improves 11 points (PAC25 10 points) which is in accord with the claimed improvement of the PFSA matched-longitudinal sample. Note in particular the results for NCE (normal curve equivalent) the metric (not necessarily a good one) in which PFSA chooses to express the year-to-year comparison for the longitudinally matched student scores; the cross-sectional NCE shows a 5.7 point gain year-to-year. Thus the designation of Clifford as a lead example of a big

decliner from cross-sectional data in the PFSA report is not consistent with the data.

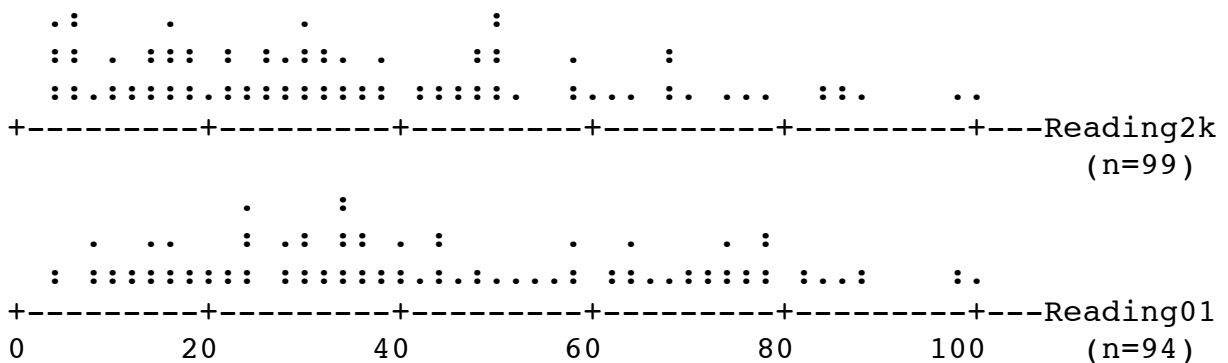
Clifford Street Elementary CDS 19647336016513

All Students in API (n2k=101, n01=94)

	2000	2001	Gain
CDEAPI	592	705	113
Reading Component API	521.7	604.3	82.6
Mean Reading NCE	40.9	46.6	5.7
Reading PAC50	26.3	37.2	10.9
Reading PAC25	62.6	72.3	9.7

One additional data analysis point is that for a school like Clifford, the PAC50 used by PFSA is not the best measure to summarize progress, because many students can make good year-to-year progress without getting anywhere near the 50th percentile. As seen from the dotplots for Stanford 9 Reading percentile rank scores, PAC25 represents a more sensitive threshold, and of course the actual API has multiple thresholds (at the quintiles). (Note: two year 2000 API students have missing Reading scores.)

Reading Percentile Rank Scores



LAUSD longitudinally-matched results.

From the LAUSD reports, based on the longitudinally matched samples with the same students in 2000 and 2001, the longitudinal gain results for Clifford can be assembled. The year-to-year improvements are for a group of students at a grade level in year 2000 followed up through the next grade in year 2001. Test scores are stated in the NCE metric, the important property here being that a student improving from the 15th percentile to the 22nd percentile will show a large year-to-year improvement, whereas that improvement would have no effect on PAC50 (or PAC25), the measure PFSA uses for cross-sectional data. (Of course, this same improvement would make

a substantial contribution to the API which PFSA claim to be analyzing.)

Clifford, Longitudinally-matched LAUSD Reading (NCE scores)

Grade	N	2000	2001	Gain
2	26	46.7	53.7	7.0
3	17	38.9	39.6	0.6
4	26	42.9	48.4	5.5
5	18	45.6	43.6	-2.1
ALL	87	43.8	47.2	3.4

The grade 3 row shows the 17 students in Clifford grade 2 in year 2000 who could be tracked within LAUSD for grade 3 scores on those same students in 2001. It is these 17 students that supply all the data used by PFSA ; the listed longitudinal gain of .6 in PFSA Figure 3 does badly understate the improvement of the longitudinally matched set of students. Over grades 2-5 for the 87 longitudinally matched Clifford students the Reading gain in the NCE metric is 3.4 points, which is reasonably close to the gain of 5.7 points seen from year 2000 and year 2001 grades 2-5 cross-sections. Interestingly, PFSA introduces the Clifford example to show cross-sectional data vastly understating the longitudinal gain, yet a reasonable comparison shows that for Clifford the NCE gain seen from comparing the 2000 and 2001 cross-sections exceeds the gain seen from the longitudinally matched sample (5.7 vs. 3.4).

Moreover, a comparison more relevant to the API would be a school-level composite across the Stanford 9 battery. For the the longitudinally matched Clifford students, the composite using the same test weights as the API yields a year 2000 score of 45.5 and a year 2001 score 53.9 for a (NCE metric) gain of 8.4, which is reasonably consistent in magnitude to the 113 point year-to-year improvement in the reported CDE API.

B. Clover Avenue Elementary, 19647336016521

Clover is also identified by PFSA Figure 3 as having strong decline, -10 PAC points, between 2000 and 2001. Examination of this school's API report shows Clover is a very high achieving school with year 2000 API of 853 and year 2001 growth API of 877, an increase of 24 API points (even though many students are topping out on the Stanford 9). To see where the -10 number for Clover in PFSA Figure 3 comes from, examine the STAR reading reports for Clover (PAC75 shown in addition to PAC50 for this high scoring school).

Grade	SAT-9 Reading, STAR Report					
	n	Year 2000		n	Year 2001	
		PAC50	PAC75		PAC50	PAC75
2	76	86	51	51	86	59
3	65	85	55	77	75	49
4	78	79	64	70	81	56
5	75	68	41	69	77	48

As discussed above PFSA takes Reading scores for a single grade in 2000, grade 3, and compares with the same grade in 2001. That cross-sectional comparison, which corresponds to nothing that is done in the API, would show no decline for Clover in grades 2,4, or 5. But comparing the grade 3 year 2000 Reading PAC50 versus grade 3 year 2001 Reading PAC50 produces $75 - 85 = -10$, and thus Clover is labeled a cross-sectional decliner.

Once again, the school level summaries based on the STAR Reading data do not show the decline indicated by PFSA. The school-wide year 2000 PAC50 for Reading is 79.25 and the year 2001 PAC50 for Reading is 79.4. An alternative school-level summary, perhaps more appropriate for Clover, PAC75, yields school-wide year 2000 PAC75 of 53.1 and the year 2001 PAC75 for Reading of 52.4. Thus from STAR Reading, cross-sectional scores for Clover are essentially the same for 2000, 2001.

To get closer to what is done in the API, consider results for the subsets of students included in the 2000 and 2001 API. On every school-wide measure,

Clover, contrary to the PFSA classification, does not decline from year 2000 to 2001. Instead of the 10 point decline in PAC50 claimed by PFSA (Fig 3), for the cohorts used in the yearly API, the Reading PAC50 measure actually improves 2 points and Reading component of the API improves 25 points in accord with the claimed improvement of the PFSA matched-longitudinal sample. Note in particular the results for NCE (normal curve equivalent) the metric (not necessarily a good one) in which PFSA chooses to express the year-to-year comparison for the longitudinally matched student scores which shows a 2 point gain year-to-year.

Clover Avenue Elementary, CDS 19647336016521
 All Students in API (n2k=268, n01=259)

	2000	2001	Gain
CDEAPI	853	877	24
Reading Component API	833.9	858.5	24.6
Mean Reading NCE	64.7	66.7	2.0
Reading PAC50	79.9	81.9	2
Reading PAC75	56	56	0

LAUSD longitudinally-matched results.

From the LAUSD reports, longitudinally matched samples with same students in 2000 and 2001, the longitudinal gain results for Clover can be assembled. The year-to-year improvements are for a group of students at a grade level in year 2000 followed up through the next grade in year 2001. Test scores are stated in the NCE metric, the important property here being that a student improving from the 77th percentile to the 87th percentile will show a large year-to-year improvement, whereas that improvement would have no effect on PAC50 (or PAC75) the measure PFSA uses for cross-sectional data. (Of course, this same improvement would make a contribution to the API which PFSA claim to be analyzing.)

Clover, Longitudinally-matched LAUSD Reading (NCE scores)

Grade	N	2000	2001	Gain
2	48	69.6	68.6	-1.0
3	66	65.5	67.7	2.2
4	55	67.2	72.1	4.9
5	59	65.5	65.3	-0.2
ALL	228	66.8	68.3	1.6

The grade 3 row shows the 66 students in Clover grade 2 in year 2000 who could be tracked within LAUSD for grade 3 scores on those same students in 2001. The listed longitudinal gain of 2.2 in PFSA Figure 3 from these 66 students overstates the improvement of the longitudinally matched set of students. Over grades 2-5 for the 228 longitudinally matched Clifford students the Reading gain in the NCE metric is 1.6 points, rather close to the gain of 2.0 points from comparing the year 2000 grades 2-5 cross-section with the year 2001 grades 2-5 cross-section. Moreover, a comparison more relevant to the API would be a school-level composite across the Stanford 9 battery. For the the longitudinally matched Clover students, the composite using the same test weights as the API yields a year 2000 score of 69.6 and year 2001 score 70.8 for a (NCE metric) gain of 1.2 which is little smaller in magnitude than the 24 point year-to-year improvement in the reported CDE API. But again in the Clover example, no discord is seen between the indications from the current practice of repeated cross-sections and the determination from the "better" longitudinally matched data.

C. Kittridge Street Elementary, 19647336017735

Kittridge is identified by PFSA Figure 3 as having cross-sectional gain but longitudinal decline. Kittridge is a relatively low-scoring school with almost all of its students classified as Socioeconomically Disadvantaged. Year 2000 API is 500 and year 2001 growth API is 556, an increase of 56 API points. From the STAR reading reports for Kittridge (PAC25 shown in addition to PAC50 for this lower scoring school), the grade 2-5 results are shown.

Grade	n	SAT-9 Reading, STAR Report		n	Year 2001	
		Year 2000			PAC50	PAC25
		PAC50	PAC25		PAC50	PAC25
2	166	26	53	193	30	58
3	154	14	42	157	18	54
4	173	17	44	143	15	50
5	135	15	35	159	25	54

As discussed above PFSA takes a single grade, grade 3, in 2000 and compares with the same grade in 2001; the grade 3 year 2000 Reading PAC50 is 14 and grade 3 year 2001 Reading PAC50 is 18, yielding $18 - 14 = 4$. Thus Kittridge is labeled a cross-sectional improver in PFSA.

The school level summaries based on the STAR Reading data do show year-to-year improvement. The school-wide year 2000 PAC50 for Reading is 18.15, and the year 2001 PAC50 for Reading is 22.55. An alternative school-level summary, perhaps more appropriate for Kittridge, PAC25, yields school-wide year 2000 PAC25 of 43.95 and the year 2001 PAC25 for Reading of 54.45.

To get closer to what is done in the API, consider results for the subsets of students included in the 2000 and 2001 API. On every school-wide measure, Kittridge improves from year 2000 to 2001. Note in particular the results for NCE, the metric in which PFSA chooses to express the year-to-year comparison for the longitudinally matched student scores, which shows a 4.7 point gain year-to-year.

Kittridge Street Elementary, CDS 19647336017735
 All Students in API (n2k=566, n01=601)

	2000	2001	Gain
CDEAPI	500	556	56
Reading Component API	425.6	491.2	65.6
Mean Reading NCE	33.9	38.6	4.7
Reading PAC50	16.4	22.2	5.8
Reading PAC25	43.9	55.3	11.4

LAUSD longitudinally-matched results.

From the LAUSD reports, the longitudinally matched samples with same students in 2000 and 2001, the longitudinal gain results for Kittridge can be assembled. The year-to-year improvements are for a group of students at a grade level in year 2000 followed up through the next grade in year 2001. Test scores are stated in the NCE metric.

Kittridge, Longitudinally-matched LAUSD Reading (NCE scores)

Grade	N	2000	2001	Gain
2	170	38.3	41.3	3
3	131	41.3	39.1	-2.2
4	126	34.3	37.4	3
5	144	34.6	39.7	5.1
ALL	571	37.2	39.5	2.4

The grade 3 row shows the 131 students in Kittridge grade 2 in year 2000 who could be tracked within LAUSD for grade 3 scores on those same students in 2001. The listed longitudinal decline, -2.2, in PFSA Figure 3 from these 131 students is in contrast to the improvement of the other three grade cohorts in the LAUSD data. Over grades 2-5 for the 571 longitudinally matched Kittridge students the Reading gain in the NCE metric is 2.4 points, and therefore the PFSA classification of Kittridge as a decliner for the longitudinal data is not persuasive. In fact, the grade 2-5 NCE gain for the longitudinally matched Kittridge sample is half-again as large as the 1.6 result for Clover, which PFSA considered to be a strong improver.

Moreover, a comparison more relevant to the API would be a school-level composite across the Stanford 9 battery. For the longitudinally matched Kittridge students, the composite using the same test weights as the API yields a year 2000 score of 40.81 and year 2001 score 43.46 for a (NCE metric) gain of 2.65. Once again, the Kittridge example displays no discord between the indications from the current practice of repeated cross-sections and the determination from the "better" longitudinally matched data.

D. Lanai Road Elementary, 19647336017768

Lanai is identified by PFSA Figure 3 as having cross-sectional gain but longitudinal decline. From CDE reports Lanai has year 2000 API of 548 and year 2001 growth API of 636, an increase of 88 API points. From the STAR reading reports for Lanai (PAC25 shown in addition to PAC50 for this lower scoring school), the grade 2-5 results are shown.

Grade	SAT-9 Reading, STAR Report					
	Year 2000			Year 2001		
	n	PAC50	PAC25	n	PAC50	PAC25
2	61	44	64	65	45	68
3	78	17	40	72	44	65
4	97	20	46	71	28	58
5	77	34	61	94	40	69

As discussed above, PFSA takes a single grade, grade 3, in 2000 and compares with the same grade in 2001. For grades 2, 4, and 5 that cross-sectional comparison would show modest to good improvement in Reading PAC50. But comparing the grade 3 year 2000 Reading PAC50 versus grade 3 year 2001 PAC50 produces $44 - 17 = 27$, and thus for PFSA, Lanai appears to display remarkable cross-sectional improvement.

The school level summaries based on the STAR Reading data are more reasonable measures of year-to-year improvement, which, as also indicated by the API, is strong. The school-wide year 2000 PAC50 for Reading is 27.16, and the year 2001 PAC50 for Reading is 39.4, a 12 point PAC50 gain. An alternative school-level summary, perhaps more appropriate for Lanai, PAC25, yields school-wide year 2000 PAC25 of 51.76 and the year 2001 PAC25 for Reading of 65.23.

To get closer to what is done in the API, consider results for the subsets of students included in the 2000 and 2001 API. On every school-wide measure, Lanai improves from year 2000 to 2001. Note in particular that the results for the NCE metric shows a 3.9 point Reading gain year-to-year, less than that for Clifford or Kittridge.

Lanai Road Elementary, CDS 19647336017768
 All Students in API (n2k=317, n01=296)

	2000	2001	Gain
CDEAPI	548	636	88
Reading Component API	517.1	583.2	66.1
Mean Reading NCE	40.3	44.2	3.9
Reading PAC50	27.2	39.9	12.7
Reading PAC25	52.5	65.7	13.2

LAUSD longitudinally-matched results.

From the LAUSD longitudinally matched samples (same students in 2000 and 2001), the year-to-year improvements are for a group of students at a grade level in year 2000 followed up through the next grade in year 2001. Test scores are stated in the NCE metric.

Lanai, Longitudinally-matched LAUSD Reading (NCE scores)

Grade	N	2000	2001	Gain
2	54	42.9	47	4.1
3	56	48.9	48.5	-0.3
4	62	34.7	39.1	4.4
5	81	37.8	47.8	10.0
ALL	253	40.6	45.6	5.1

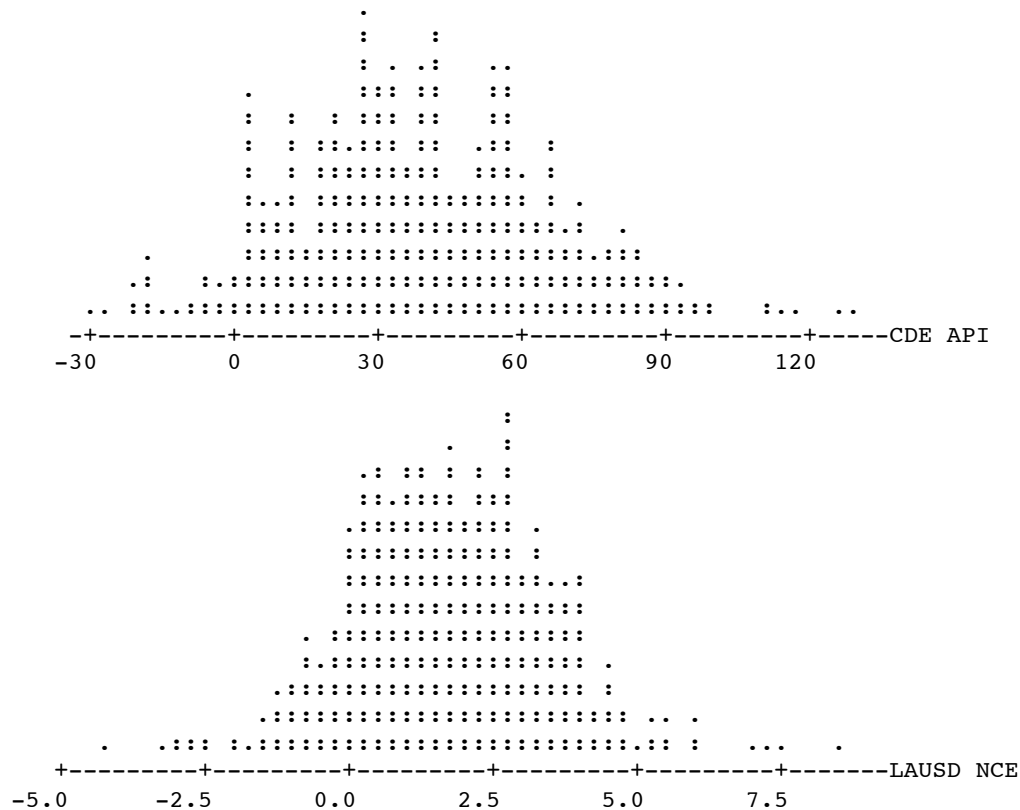
The grade 3 row shows the 56 students in Lanai grade 2 in year 2000 who could be tracked within LAUSD for grade 3 scores on those same students in 2001. The listed longitudinal decline, -0.3, in PFSA Figure 3 from these 56 students is in contrast to the improvement of the other three grade cohorts in the LAUSD data. Over grades 2-5 for the 253 longitudinally matched Lanai students the Reading gain in the NCE metric is 5.1 points, making Lanai the largest improver of all four exemplar schools. Therefore the PFSA classification of Lanai as a decliner for the longitudinal data is contrary to the data.

Moreover, a comparison more relevant to the API would be a school-level composite across the Stanford 9 battery. For the longitudinally matched Lanai students, the composite using the same test weights as the API yields a year 2000 score of 42.81 and year 2001 score 49.82 for a (NCE metric) gain of 7.01. Once more, in the Lanai example, no discord is seen between the indications from the current practice of repeated cross-sections and the determination from the "better" longitudinally matched data.

3. Comparing Cross-sectional and Longitudinal Results for the Collection of LAUSD Elementary Schools

The PFSA main display, their Figure 4 ("Percentage Agreement on School Progress Between the California State API and Longitudinal Data from the LAUSD Schools"), purports to compare the CDE API with a school summary based on the matched longitudinal sample. The preceding examination of the PFSA school examples has demonstrated the misleading nature of that labeling and presentation (a single subject on a single grade with a different measure and a different sample). To supplement the reexamination of the four PFSA example schools, here we present the comparison that PFSA claimed to have done. This collection of 403 LAUSD elementary schools has API scores in 2000 and 2001 and matched longitudinal scores in the LAUSD data files (3 of the 406 LAUSD schools do not have reported API for 2001).

For each LAUSD school, the API measure is the improvement in the CDE reported API scores between 2000 and 2001. For the matched longitudinal sample a composite across tests using the same test weights as the API (i.e., $.3 * \text{meanNCEReading} + .4 * \text{meanNCEMath} + .15 * \text{meanNCELanguage} + .15 * \text{meanNCESpell}$) produces school level scores for year 2000 and for year 2001, from which the measure of improvement in the NCE scale is computed. Dotplots and descriptive statistics for the two improvement measures are shown below.



Descriptive statistics for improvement

Improvement	N	Median	Q1	Q3	Minimum	Maximum
CDE API	403	37.87	20.75	57.38	-28.94	128.62
LAUSD NCE	403	1.74	0.48	2.87	-4.34	8.38

The scatterplot in Figure 1 displays the correspondence between the two scores, which is also summarized in the table below. The question raised in the PFSA report is the prevalence of contrary indications on school progress (positive vs. negative progress) between the currently used API and a school measure obtained from longitudinally matched individual data. The PFSA report claimed a 40% discord rate; the display below indicates 8% (5 + 26 out of 403). The three-category division used in the table below, and also shown by the superimposed lines in Figure 1, creates a narrow zone around 0 (one-half point for NCE and 5 points for API) which for these purposes is regarded as zero improvement. (As can be seen in Figure 1, making this middle category smaller increases the discord rate slightly.)

CDE API	NCE-LAUSD matched sample composite		
	Gain < -.5	Gain in [-.5, .5]	Gain > .5
Gain < -5	6 (1.49%)	5 (1.24%)	5 (1.24%)
Gain in [-5, 5]	5 (1.24%)	11 (2.73%)	13 (3.23%)
Gain > 5.0	26 (6.45%)	49 (12.16%)	283 (70.22%)

Because the students in the matched longitudinal sample are not the same as those in the API, disagreements for some school scores are inevitable. On the average, the size of matched-longitudinal sample is about 10% less than that used for the school's API score, but those numbers vary widely over the collection of schools. The size of the matched longitudinal sample ranges from half as large as the number of students in the API to one-and-one-half times as large. To illustrate, consider the example of Fletcher Drive Elementary (CDS 19647336017115). The 2000 and 2001 STAR reports show 644 and 731 students for grades 2-5. The school API reports show 410 students in 2000 and 474 students in 2001, with more than 1/3 of the STAR students excised each year for school mobility as required by PSAA legislation. The LAUSD matched longitudinal sample has 570 Reading and 604 Math scores, thus including many students not included in the API. Which set of students is the most appropriate for school accountability is unclear, but clearly, school summaries based on rather different sets of students will show some discord.

CDE API Gain

CDE API Gain vs LAUSD NCE Gain

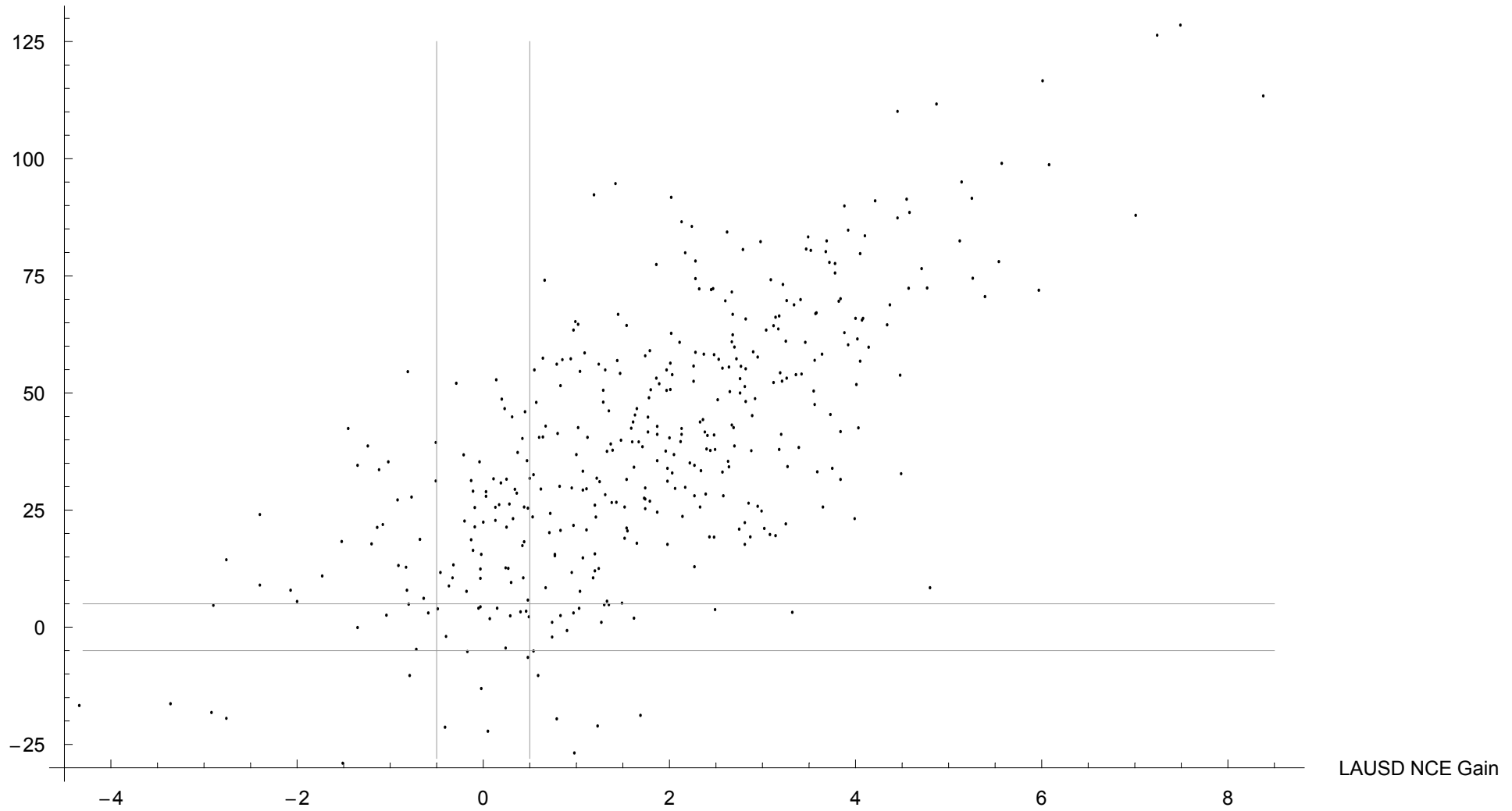


Figure 1. Plot of improvement from 2000 to 2001 for 403 LAUSD elementary schools on API (vertical axis) using the students included in the CDE reported API and on mean NCE (horizontal axis) for the students included in the matched longitudinal sample.

4. Conclusion.

Clearly, these extended examples demonstrate that the empirical claims in the PFSA report of deficiencies in the current API are not credible.

Specifically, when the data for the PFSA example schools are examined in any reasonable manner, the claimed discrepancies between the cross-sectional and longitudinal results disappear. One summary table is given below: the actual reported CDE API compared to a NCE composite score computed from the LAUSD longitudinal matched sample (Stanford 9 tests Reading, Math, Language, Spelling weighted the same as in the API).

Summary Table

School	CDE API			Longitudinal Matched NCE		
	2000	2001	Gain	2000	2001	Gain
Clifford	592	705	113	45.5	53.9	8.4
Clover	853	877	24	69.6	70.8	1.2
Kittridge	500	556	56	40.8	43.5	2.7
Lanai	548	636	88	42.8	49.8	7.0

To be sure, noone can argue against better data, although the cost-benefit issues may be serious. By the same token, noone can argue against competent data analysis and presentation. And, on this, the PFSA report is seen to fall short.

References

2000 to 2001 Stanford 9 Matched Individual Student NCE Gains by District, School, and Grade Level. Los Angeles Unified School District Program Evaluation and Research Branch, October 2001.

http://www.lausd.k12.ca.us/lausd/offices/perb/files/pau/SAT901/NCEGainReport2001_101101.pdf

A Better Data System for California. Oakland CA: Public Forum on School Accountability, P. Daro, Director, April 2003.

EdSource Report, May 2003. EdSource's 26th Annual Forum—California School Reforms: Show Me the Data!

http://www.edsource.org/pdf/Forum03report_Final.pdf

Linn, R. L., & Haug, C. (2002). Stability of School-Building Accountability Scores and Gains. *Educational Evaluation and Policy Analysis*, 24, 29-36.

Rogosa, D. R. Confusions about Consistency in Improvement. Stanford University, CRESST, June 2003. Available from

<http://www-stat.stanford.edu/~rag/api/consist.pdf>