Student Progress in California Charter Schools, 1999-2002

David Rogosa
Stanford University
June 2003

Acknowledgements

Thanks to Ghassan Ghandour for the usual computing wizardry.

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002 and Award Number R305B960002-01 as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education. Additional support was provided by the California Department of Education, Policy and Evaluation Division.

Student Progress in California Charter Schools, 1999-2002

David Rogosa
Stanford University
June 2003

1. Background and Summary

Start with a bit of background on these analyses. In March 2002 Slovacek, Kunnan, and Kim (hereafter SKK) compared API scores for charter and non-charter schools over the period 1999-2001 (with an additional emphasis on socioeconomic factors). Rogosa (2002) refuted the SKK analyses, showing that mistakes in the identification of schools and in the assembly of school data, plus flaws in their data analysis approaches, render the SKK conclusions incorrect. Rogosa (2002) presented useful data analysis approaches for the student test data, applied to the comparison of student scores in charter and non-charters over the 1999-2001 period. Recently, Raymond (2003) repeated the SKK school-level comparisons of charter versus non-charter schools for 1999-2001, unfortunately also replicating most of the methodological shortcomings in SKK (and adding a few new problems).

The present analyses use Stanford 9 scores over the four-year period 1999-2002, thereby extending the analyses in Rogosa (2002) with another year of test data. Recognizing that different readers will have different appetites for details, below are some bottom-line summaries of the "horse-race" results. Further description (especially description by grade-level and bands of grade levels) and justification for these numbers are developed at length in the body of the report.

Improvement of all Students		
	Charter Students	Non-charter students
1999 API	597.28	621.13
2000 API	626.7	649.13
2001 API	633.98	661.89
2002 API	648.55	672.75
'99-02 improvement	51.27	51.66
Improvement of Socioeconomically Disadvantaged (SD) Students		
	Charter Students	Non-charter students
1999 API	485.02	503.56
2000 API	517.36	538.31
2001 API	533.95	557.23
2002 API	555.7	576.08
'99-02 improvement	70.68	72.60
Improvement of Socioeconomically Disadvantaged (SD) Students in High SD Schools		
	Charter Students	Non-charter students
1999 API	457.51	481.81
2000 API	492.6	517.27
2001 API	510.73	539.79
2002 API	540.5	562.21
'99-02 improvement	82.99	80.40

The 1999-2001 results for student progress in Rogosa (2002) indicated a consistent 4 point advantage to non-charter schools (opposite the SKK and Raymond findings) over the three groups of students for grade span 2-11 (which can be replicated by comparing 1999 and 2001 scores above). That same differential in student progress was also seen in the grades 2-6 subsets for the All Student and SD student groups for 1999-2001. The new results above for the 1999-2002 period show even less differences between the student progress in charter and non-charter schools. (sidenote: I use the term "non-charter", where others use terms like "conventional" or "traditional" in describing the bulk of schools.)

Perhaps the most misunderstood item in the various charter, non-charter comparisons is the metric of the California API score, especially that differences of 2 or 4 API points are not educationally important. One explanation in Rogosa (2002) was that: "An appropriate interpretation of a 4 API point difference is: every other student getting an additional item correct on each Stanford 9 test. Thus, in educational terms 4 API points is not that large an amount." Another version of a four point difference would be every student getting an additional item correct on the Stanford 9 mathematics test (see Rogosa 2000). Data analyses and interpretations for the API scale can be found in the "Interpretive Notes" series which are available from the California Department of Education API Research page (http://www.cde.ca.gov/psaa/apiresearch.htm).

The body of this report pursues various additional comparisons of student progress in charter and non-charter schools and explains some of the myriad deficiencies of the other analyses. Readers interested in the results for charter schools should focus on Section 3 (especially. Exhibits 1-4 and Figures 1-2). Section 4, which focuses on methodological issues (especially the data analysis and statistical shortcomings in the Raymond and SKK analyses), will be of more interest to those conducting and reviewing research. The content of Section 4 serves to confirm the maxim--"fancier is rarely, if ever, better"--and to urge researchers to embrace solid, common-sense data analysis.

2. Student Data and School Classification

Charter School Identification.
The 93 charter schools (CDS code listing in Appendix A) are the same schools used in the Rogosa (2002) 1999-2001 analyses; Rogosa (2002, sec. 2) contains further details on the assembly of this list. All 93 charters have student test data for 1999, 2001, and 2002. Five of the schools have missing scores in 2000. Three of the 93 schools did not have reported API in 2002 because of participation or demographic criteria, but the student scores were included in this analysis (as these analyses do not evaluate individual schools).

Non-Charter School Identification.
Start with a set of 7672 California Schools having some API data in the 1999-2002 period. From the CDE California School Directory (located at www.cde.ca.gov/schooldir) obtain (in spring 2002) a list of CDS codes for 372 active charter schools plus 21 charter schools listed as merged/closed. Remove these 393 CDS codes from the set of 7672 California schools, and select from these remaining (non-charter) schools those having API scores for the 1999-2002 period. This process produces a set of 6584 non-charter schools having API data in at least 1999, 2001, 2002 (77 of these having missing API data in 2000).

Designation of SD (Socioeconomically Disadvantaged) students and High SD Schools.
In California's accountability system a student is classified as Socioeconomically Disadvantaged (SD) if either:
1) Neither of the pupil's parents has received a high school diploma
OR
2) The pupil participates in the National School Lunch Program (NSLP), free or reduced price lunch program.

The High SD subset of schools is composed of schools having at least 50% of students designated as SD in at least 3 years of the 1999-2002 period. As a result, 41 of the 93 charter schools are classified HighSD, and 3135 of the 6584 non-charter schools are classified HighSD.

Individual Student Data.
For each of students contained in the analysis, the individual-level data consist of the year, grade level, Stanford 9 test battery scores, and SD designation. For any specified collection of students an API index score can be computed from the Stanford 9 battery scores. In particular, for 2002 the same form API index was constructed from the Stanford 9 portion of the STAR testing as was employed in 1999, 2000, and 2001(growth) indices. School-level data attached to each student include the charter/non-charter and HighSD designations. For grades 2-11 the non-charter school data contains scores for more than 350,000 students each year and the charter school data contains scores for about 40,000 students each year.

Even though the test data is identified with a grade level in a school (and with the individual's characteristics) the individual test data are not matched longitudinally. That is, individuals cannot be traced over years of testing. Consequently, the kind of improvement that can be calculated is at the school and grade level either as successive cross sections (such as

second-graders in 1999 versus second-graders in 2002 as in Exhibits 1-3) or as a rough longitudinal cohort (such as second-graders in 1999 versus fifth-graders in 2002 as in Exhibit 4). And thus the term "student progress" used throughout is a shorthand for improvement of successive cross-sections (implemented in Exhibits 1-3) or, as implemented in Exhibit 4, improvement for approximate longitudinal cohorts.

The distribution of student data across grade levels does show some differences between the charter and non-charter groups, with a greater proportion of charter students in the lower grades and a smaller proportion in the upper grades. This difference is consequential because analyses in the Interpretive Notes series have shown API improvement is much larger in the lower grade levels, and thus a sample overweighted in the lower grades would be advantaged in a comparison of student progress. The detailed analyses to follow display the charter/non-charter comparisons over specified grade ranges (e.g. grades 2-6, 2-8) and grade-by-grade to avoid this confounding. The collections of grades (as in the Create One Large School tables) are useful (even necessary) to smooth uneven results in the Charter Schools over grade levels. Schooltype (Elementary, Middle, High) is not an adequate proxy for grade level as Schooltype designations are fluid.

Average Proportion of Students in Each Grade

Grade	charter	non-charter
2	0.130	0.107
3	0.132	0.109
4	0.134	0.109
5	0.128	0.109
6	0.102	0.105
7	0.094	0.102
8	0.094	0.101
9	0.070	0.092
10	0.064	0.090
11	0.051	0.077

The proportion of SD students does not differ much between the charter and non-charter schools; overall charters have 49% of students designated SD and noncharters have 48% of students SD. Below is the proportion of SD students by grade-level.

Average Proportion of SD Students in Each Grade

Grade	charter	non-charter
2	0.584	0.565
3	0.575	0.556
4	0.564	0.544
5	0.567	0.532
6	0.495	0.507
7	0.429	0.470
8	0.380	0.439
9	0.418	0.396
10	0.302	0.351
11	0.266	0.320

3. Description of Student Improvement in Charter and non-Charter Schools

The main comparisons of student progress in charter and non-charter schools are displayed in Exhibits 1-3 and Figures 1 and 2. Each Exhibit displays results for different pools of students:
Exhibit 1, all students in the 93 charter and the 6584 non-charter schools,
Exhibit 2, Socioeconomically Disadvantaged (SD) students in the 93 charter and the 6584 non-charter schools,
Exhibit 3, SD students in the subset of High SD schools (41 charter schools and 3135 non-charter schools).
Exhibits 1-3 follow the progression of conditioning: start with all students in Exhibit 1, then refine to SD students in Exhibit 2, then refine further to SD students in the subset of schools that are classified as High SD in Exhibit 3. This successive conditioning data analysis strategy represents an implementation of the "disaggregation of results" mantra commonly associated with No Child Left Behind.

Each Exhibit contains 3 tables. The bottom table "Complete Grade-by-Grade Display" displays an API score at each grade level and at each year for students in charter and non-charter schools. Also, the right-hand columns show the number of students at each grade level for each year. For example, the "Complete Grade-by-Grade Display" in Exhibit 1 shows a Charter Student Grade 2 1999 API score of 585 (computed using the data from the 5319 Grade 2 charter school students as one large school). The top table "Grade-by-Grade Improvement" is derived from the "Complete Grade-by- Grade Display" by taking the difference between the 2002 and 1999 API scores. The middle table "Create One Large School Table" is my best attempt at summary. The grade-by-grade displays are aggregated to the stated grade interval: e.g., all students with grades 2-11, or some version of Elementary School students using grade intervals of 2-6 or 2-8.

Why display grade-by-grade results or collections of grades (as in the Create One Large School tables)? Two interrelated reasons: (i) that charter schools may have an unconventional grade span or distribution, and so breaking out grade levels allows clearer comparisons with non-charter schools; (ii) analyses in the Interpretive Notes series have shown much larger API improvement for the lower grade levels. Therefore, the combinations over grade levels in the Create One Large School tables are more useful summaries.But the grade-by-grade numbers are interesting, especially if one is careful not to overinterpret results based on relatively small amounts of data. For example, it's important to note that for the subset of charter schools classified as High SD (Exhibit 3) only 2 schools contribute grade 10 and 11 students (98% of grade 11 from one school), 3 schools contribute grade 9 students, and although 11 schools contribute grade 8 students nearly 90% of those grade 8 students are from three schools.

Summary of results, Exhibits 1-3.
Grade-by-Grade Improvement is strong in the lower grades, but for charter students improvement is uneven across upper grades in each of the breakdowns. The comparisons using a band of grades (Create One Large School) show charter and non-charter students making comparable (sometimes close to identical) improvement over the four year period. The All Students

(Exhibit 1) grade 2-11 comparison shows identical improvement for charter and non-charter students; the charter students show slightly greater improvement in the lower grades, which is reversed in the upper grades. For SD students (Exhibit 2) non-charter students make slightly larger improvement over each of the grade bands, but when only HighSD schools are included (Exhibit 3) the slight advantage in improvement is for the SD charter school students (seen in the lower grades).

Insert Exhibits 1-3

Figures 1 and 2 seek to provide some additional perspective by displaying the considerable improvement over the 4 years by both charter and non-charter students, especially in comparison to the (very small) differentials in improvement in comparison to the large differences in the scores. The API scores for the three groups--All Students, SD Students, SD Students in High SD Schools--conform in descending order to the order of listing. And within each grouping of students, charter school students (in red) have lower scores than non- charter students (in black). (A visual challenge in Figure 1 is produced because the scores for SD students in charter schools are nearly identical to SD students in High SD non-charter schools.) Looking at Figures 1 and 2, especially over the 1999-2001 period, it is hard to comprehend the assertion by Margaret Raymond in the Los Angeles Times (June 18, 2003): "If these trends continue, charter schools in another few years will be not only at the level of traditional schools but exceeding them."

Insert Figures 1 and 2

Comparing Cohorts.
Exhibits 1-3 provide comparisons of repeated cross sections (e.g. 3rd graders in 1999 compared with 3rd graders in 2002). Exhibit 4 augments those comparisons by tracing (rough) cohorts over time. Cohorts (and their sizes) can be traced by following diagonals in the Complete Grade-by-Grade Display at the bottom of each Exhibit. Exhibit 4 organizes the grade-by-year results into cohort summaries. The upper half shows results for all students and the lower half for SD students. The "grade-by-grade cohort improvement" displays the difference between 2002 API and 1999 API scores for the cohort beginning in 1999 in the designated grade. For example, charter students in grade 2 in 1999 are in grade 5 in 2002, and the table value for all charter students of 52.32 for 1999 Grade 2 is obtained from Exhibit 1 as 637.32 - 585. As with Exhibits 1-3 the charter students show irregular improvement in the upper grades, and it's important to keep in mind that a small number of schools are contributing those students. Perhaps of most use are the "combined cohort" comparisons; these include students in grades 2-5 in 1999 and grades 5-8 in 2002 (and grades 2-3 in 1999, 5-6 in 2002). For both the all students and SD students tables, the combined cohorts constructed for the charter students show overall improvement 1 to 4 points larger than for the non-charter students. The pattern of results for these artificial cohorts with all students mirror the results (Exhibit 1) seen for the successive cross-sections in the lower grades. For SD students, using the cohorts rather than successive cross-sections changes the horse-race advantage slightly in favor of charters (compare with Exhibit 2).

Insert Exhibit 4

Exhibit 1
All Students in the Charter and Non-charter Schools

Grade-by-Grade 1999-2002 Improvement Table

Grade	Charter Students	Non-charter students
2	89.02	77.75
3	81.89	83.62
4	91.76	83.75
5	59.57	67.88
6	52.12	54.88
7	7.62	40.88
8	5.5	29.
9	4.75	24.38
10	19.12	16.38
11	23.88	17.38

Create One Large School Table

Grades Included		1999 API	2001 API	2002 API	Improvement '99-02
2-6	Charter Students	587.92	641.28	664.15	76.22
	Non-charter students	620.22	676.71	693.83	73.61
2-8	Charter Students	597.42	637.42	657.14	59.72
	Non-charter students	622.36	671.49	685.27	62.92
9-11	Charter Students	596.63	618.93	611.5	14.86
	Non-charter students	617.32	634.48	637.09	19.77
2-11	Charter Students	597.28	633.98	648.55	51.27
	Non-charter students	621.08	661.86	672.75	51.66

Complete Grade-by-Grade Display

Charter Students: API Grade-by-Grade / Number of Students

Grade	API_99	API_2k	API_01	API_02	N_99	N_2k	N_01	N_02
2	585	645	649.75	674.02	5319	4788	5271	5505
3	590.75	624.625	658.625	672.64	5482	5091	5198	5428
4	572.875	612	631.5	664.64	5224	5250	5382	5581
5	577.75	609.875	619.375	637.32	5065	4828	5207	5442
6	621.5	641.75	649	673.625	3877	3876	4125	4510
7	629.75	643.75	627.375	637.375	3850	3139	3891	4244
8	627.375	642.125	622.375	632.875	3762	3194	3864	4181
9	592.625	631.5	612.375	597.375	2655	2358	3041	3230
10	575.5	584.875	608.25	594.625	2672	2365	2517	2656
11	628.625	630.625	642.625	652.5	2098	1955	1976	2205

Non-Charter Students: API Grade-by-Grade / Number of Students

Grade	API_99	API_2k	API_01	API_02	N_99	N_2k	N_01	N_02
2	627	673.25	684.75	704.75	390420	367860	381727	388928
3	619.625	665.625	688.875	703.25	404416	381330	380120	395136
4	606	646.25	668.75	689.75	385347	385358	388093	394265
5	609.25	639	660	677.125	381478	369921	394598	406268
6	640	666.375	682.125	694.875	366632	358509	367285	400218
7	625.125	646.625	658.5	666	355017	348726	362461	382331
8	631	649.875	656.625	660	363210	347457	354721	380244
9	618.375	637.875	641.75	642.75	298473	321195	335108	352324
10	594.125	605.125	608.25	610.5	320822	310990	318280	328220
11	643.375	654.25	656.25	660.75	273481	269394	271594	284433

Exhibit 2
SD Students in the Charter and Non-charter Schools

Grade-by-Grade 1999-2002 Improvement Table

Grade	Charter Students	Non-charter students
2	113.66	104.25
3	100.63	109.81
4	112.56	106.94
5	75.2	87.94
6	57.5	70.88
7	17.94	53.06
8	0.25	38.94
9	13.06	29.88
10	16.25	19.12
11	15.12	21.88

Create One Large School Table

Grades Included		1999 API	2001 API	2002 API	Improvement '99-02
2-6	Charter Students	479.12	545.14	573.4	94.28
	Non-charter students	506.1	576.43	602.43	96.33
2-8	Charter Students	484.92	537.17	563.73	78.81
	Non-charter students	505.63	567.94	589.7	84.07
9-11	Charter Students	485.69	512.38	500.47	14.78
	Non-charter students	494.4	512.57	519.3	24.9
2-11	Charter Students	485.02	533.95	555.7	70.68
	Non-charter students	503.48	557.23	576.08	72.6

Complete Grade-by-Grade Display

Charter Students: API Grade-by-Grade / Number of Students

Grade	API_99	API_2k	API_01	API_02	N_99	N_2k	N_01	N_02
2	480	546.875	561.375	593.66	2775	2817	3261	3329
3	487.5	520.875	562.375	588.13	2813	3051	3015	3322
4	454.563	501.5	529.25	567.12	2608	3138	3125	3256
5	466.125	489.375	523.25	541.32	2532	2769	3068	3307
6	520.25	535.5	550.75	577.75	1724	1972	2120	2319
7	510.563	530.375	509.438	528.5	1341	1369	1810	1975
8	519	517.625	495.75	519.25	1111	1210	1593	1809
9	486.188	527.375	510.25	499.25	970	955	1367	1458
10	458.625	473.563	496.625	474.875	787	719	800	779
11	523.25	519.875	541.875	538.375	554	538	526	573

Non-Charter Students: API Grade-by-Grade / Number of Students

Grade	API_99	API_2k	API_01	API_02	N_99	N_2k	N_01	N_02
2	521	574.25	594.5	625.25	191276	211954	225753	234802
3	508.813	565.5	594.5	618.625	195774	218327	217645	235303
4	486.938	534.125	564.625	593.875	183144	214442	219730	228362
5	490.813	524.5	552.875	578.75	178703	200190	216276	231694
6	523.625	552.875	574.75	594.5	163225	186263	190743	218122
7	501.188	526.25	542.25	554.25	144369	170245	175628	191757
8	507.188	529.375	539.5	546.125	138144	156891	160914	179273
9	505.75	527.25	530.25	535.625	105026	132388	134546	147758
10	462.813	474.5	476.313	481.938	100232	113797	113905	121326
11	519.625	532.75	532.25	541.5	78228	89274	89002	95566

Exhibit 3
SD Students in the High SD Charter and Non-charter Schools

Grade-by-Grade 1999-2002 Improvement Table

Grade	Charter Students	Non-charter students
2	120.21	110.19
3	110.86	116.38
4	124.28	112.56
5	88.4	91.56
6	73.06	72.81
7	28.5	52.75
8	-12.06	39.
9	1.81	30.69
10	0.62	20.25
11	10.81	24.62

Create One Large School Table

Grades Included		1999 API	2001 API	2002 API	Improvement '99-02
2-6	Charter Students	452.46	526.05	558.41	105.95
	Non-charter students	483.81	556.3	585.34	101.53
2-8	Charter Students	457.96	516.95	549.28	91.32
	Non-charter students	482.91	547.48	572.24	89.34
9-11	Charter Students	453.59	450.78	458.76	5.17
	Non-charter students	475.37	493.91	501.48	26.11
2-11	Charter Students	457.51	510.73	540.5	82.99
	Non-charter students	481.81	539.79	562.21	80.4

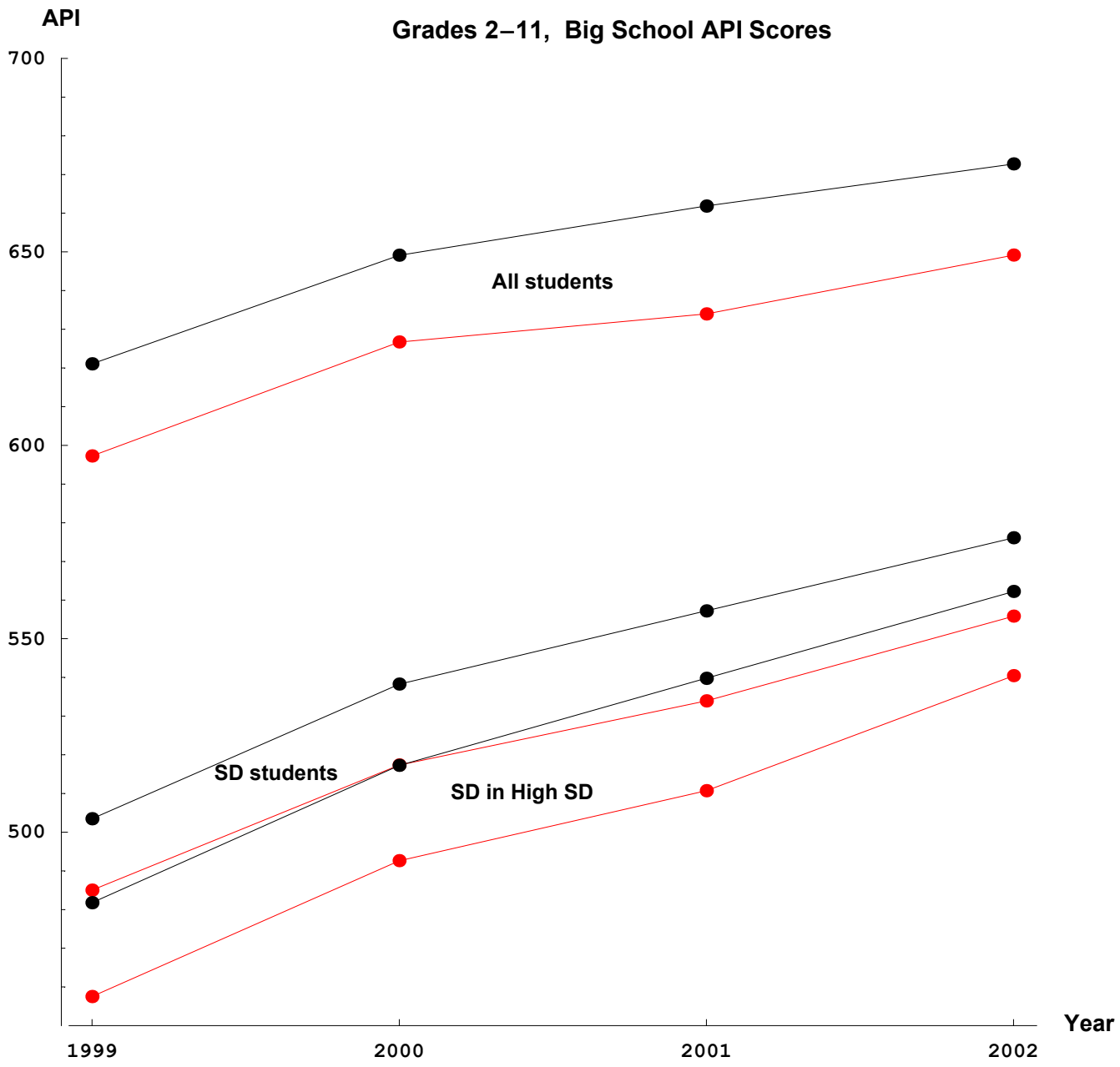
Complete Grade-by-Grade Display

Charter Students: API Grade-by-Grade / Number of Students

Grade	API_99	API_2k	API_01	API_02	N_99	N_2k	N_01	N_02
2	458.875	525.5	541.125	579.08	2339	2398	2778	2877
3	463.125	502.688	546.625	573.99	2329	2606	2573	2846
4	430.938	479.438	513.75	555.22	2142	2552	2623	2747
5	438.063	472.688	501.688	526.46	2044	2265	2548	2753
6	480.938	506.813	525.75	554	1253	1415	1610	1620
7	482.938	502	477.75	511.438	940	842	1305	1301
8	502.438	488.813	463.313	490.375	722	706	1104	1154
9	460.375	463.875	463.563	462.188	636	585	854	1040
10	413.875	428.125	413.313	414.5	418	415	389	349
11	494.563	465.875	464.563	505.375	300	308	265	255

Non-Charter Students: API Grade-by-Grade / Number of Students

Grade	API_99	API_2k	API_01	API_02	N_99	N_2k	N_01	N_02
2	501.563	554.25	577.625	611.75	157576	173256	187095	194303
3	487.125	544.5	576.5	603.5	160992	177490	177709	193220
4	463.063	510.5	543.25	575.625	149130	173670	179048	185428
5	468.563	501.375	531.875	560.125	145686	160957	175480	187665
6	499.688	527.125	549.625	572.5	125042	140018	143536	163813
7	475.625	497.688	515.75	528.375	101553	118295	123293	133588
8	483.625	503.438	515.125	522.625	98000	109583	114139	126022
9	484.063	504.938	508.375	514.75	61052	72795	74800	80491
10	443.25	451.188	458.063	463.5	55808	61274	61839	64863
11	504.625	514.875	517.625	529.25	43124	46997	47856	50256


API
Grades 2–11, Big School API Scores
700
650
600
550
500
All students
SD students
SD in High SD
1999
2000
2001
2002
Year

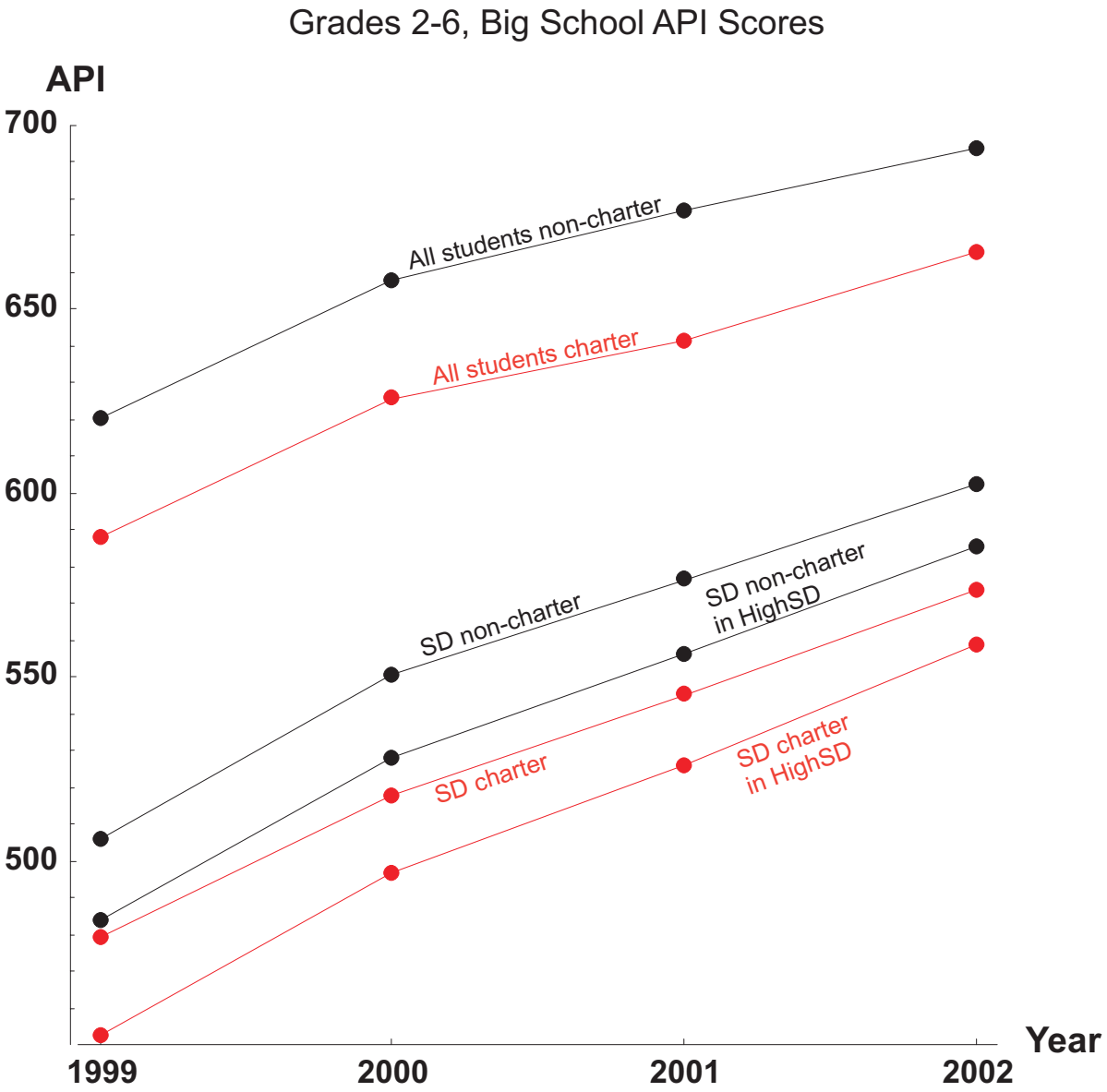

Grades 2-6, Big School API Scores
API
700
650
600
550
500
All students non-charter
All students charter
SD non-charter
SD non-charter
in HighSD
SD charter
SD charter
in HighSD
Year
1999
2000
2001
2002

Exhibit 4: Cohort Summaries

Students from Exhibit 1 (all students)

Grade-by-Grade 1999-2002 Cohort Improvement Table

1999 Grade	Charter Students	Non-charter students
2	52.32	50.12
3	82.88	75.25
4	64.5	60.
5	55.12	50.75
6	-24.12	2.75
7	-35.12	-14.62
8	25.12	29.75

API Scores Combined Cohort, 1999 Grades 2-3

Year	Charter Students	Non-charter students
1999	587.92	623.25
2000	618.22	655.89
2001	625.54	664.34
2002	653.77	685.93
Improvement '99-02	65.85	62.69

API Scores Combined Cohort, 1999 Grades 2-5

Year	Charter Students	Non-charter students
1999	581.75	615.57
2000	620.89	654.22
2001	631.12	667.26
2002	645.23	674.79
Improvement '99-02	63.48	59.22

Students from Exhibit 2 (SD students)

Grade-by-Grade 1999-2002 Cohort Improvement Table

1999 Grade	Charter Students	Non-charter students
2	61.32	57.75
3	90.25	85.69
4	73.94	67.31
5	53.12	55.31
6	-21.	12.
7	-35.69	-19.25
8	19.38	34.31

API Scores, Combined Cohort, 1999 Grades 2-3

Year	Charter Students	Non-charter students
1999	483.78	514.84
2000	511.05	549.95
2001	526.28	558.8
2002	556.34	586.39
Improvement '99-02	72.56	71.55

API Scores, Combined Cohort, 1999 Grades 2-5

Year	Charter Students	Non-charter students
1999	472.51	502.28
2000	509.97	544.4
2001	528.39	558.97
2002	543.36	570.09
Improvement '99-02	70.86	67.81

4. Methodological Concerns and Commentary

A number of enduring data analysis issues that also arise in other charter school analyses merit some explanation.

A. Mean of School Means Analyses

The Raymond and SKK analysis strategy can be summarized as

1. Obtain three years of API scores for a list of charter schools.
   1a. Compute the mean of the school scores at each of the three years (the "Mean-of-School-Means" for charter schools)
2. Obtain three years of API scores for a list of non-charter schools.
   2a. Compute the mean of the school scores at each of the three years (the "Mean-of-School-Means" for non-charter schools)
3. Compare improvement of charter and non-charter schools by subtracting the 1999 mean-of-school-means from the year 2001 mean-of-school-means.

Table 1 presents the mean-of-school means analysis for the collection of charters and non-charters used in these analyses. Exhibit 1 is the corresponding student-level analysis. For the All Schools portion of Table 1, the 1999-2002 comparison shows a 2 point advantage for non-charters, and the same 2 point advantage is seen for the subset of Elementary Schools broken out in the second part of Table 1. For the 1999-2001 period used by Raymond and SKK, the differences between charter and non-charter improvement are less than 1 point for All Schools and Elementary Schools. The small number of Middle Schools show a large non-charter advantage, which is reversed for the High Schools (pattern seen for 1999-2001 or 1999-2002 comparison).

Insert Table 1

Contrast the Table 1 indications with Exhibit 1 results for 1999-2002. In Exhibit 1, the grade 2-11 comparison was a dead heat and the grade 2-6 comparison favored charter schools by nearly 3 points. Thus the mean-of-school means analysis would appear operate to the disadvantage of charters. That impression is reversed for the high school level where the grades 9-11 comparison in Exhibit 1 shows a 5 point advantage for non-charter schools, that is opposite to the indications of Table 1 (6 point advantage for charter high schools for 1999-2002).

Furthermore, the high school comparison is one of Raymond's lead findings: "The striking finding, however, is that their [charter high schools] average improvement since 1999 is more than twice than [sic] of conventional high schools, a statistically significant difference." (p.2) Yet, there seems to be little support for this finding in the API data. Curiously, Raymond in her Graphic 6 indicates 1999-2001 improvement for charter high schools in excess of 35 points, where the actual (Table 1) mean-of-school-means improvement for 1999-2001 is 26 points.

If all schools had the same number of students, then the Table 1 and Exhibit 1 grade 2-11 results would match. In a mean-of-school-means analysis students in smaller schools contribute more to the result than students in larger schools. For the collection of Charter Schools, the displays of school size below indicates the range of school sizes are a

Table 1 Mean-of-School-Means Tables

	All Schools	
	Charter	non-Charter
	API	API
	(n)	(n)
1999	609.7	631.05
	(93)	(6584)
2000	643.9	664.78
	(88)	(6507)
2001	658.8	680.96
	(93)	(6584)
2002	671.0	694.14
	(93)	(6584)
1999-2002 Increase in Mean-of-School-Means	61.3	63.09

Table 1 continued

Mean-of-School-Means Tables By School Type 1999 (E M H)

Elementary Schools	Charter API (n)	non-Charter API (n)
1999	614.17 (68)	632.41 (4699)
2000	655.98 (66)	672.07 (4632)
2001	673.81 (68)	691.39 (4699)
2002	687.86 (68)	707.79 (4699)
1999-2002 Increase in Mean-of-School-Means	73.69	75.38

Middle Schools	Charter API (n)	non-Charter API (n)
1999	595.30 (12)	632.46 (1089)
2000	617.61 (11)	654.67 (1083)
2001	609.90 (12)	666.96 (1089)
2002	626.18 (12)	674.86 (1089)
1999-2002 Increase in Mean-of-School-Means	30.88	42.4

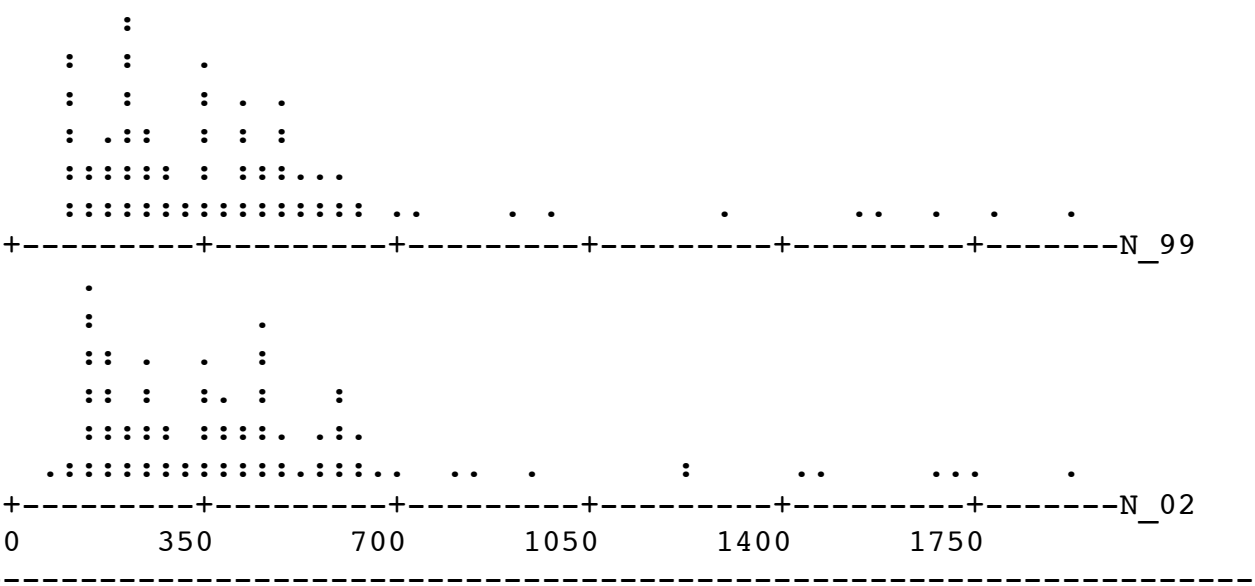
High Schools	Charter API (n)	non-Charter API (n)
1999	599.34 (13)	621.09 (796)
2000	597.56 (11)	636.01 (792)
2001	625.62 (13)	638.53 (796)
2002	624.23 (13)	640.01 (796)
1999-2002 Increase in Mean-of-School-Means	24.89	18.92

factor of 20 or more. Consequently, one student in the smallest schools contributes as much to the mean-of-school-means result as a classroom of students in the largest schools. The distortion of the mean-of-school means analysis is strongest for the set of charter high schools, for which the disparities in school size are greatest.

Descriptive Statistics: N_99, N_02 for the 93 Charter Schools

Variable	N	Median	Q1	Q3	Minimum	Maximum
N_99	93	351.0	205.0	490.0	93.0	1920.0
N_02	93	378.0	201.5	566.5	81.0	1921.0

Dotplot: N_99, N_02 for the 93 Charter Schools



An additional distortion in the mean-of-school-means analyses is that School Type (E M H) is not a perfect proxy for grade-level. Some charter schools have unconventional grade span or distribution, and so breaking the scores out by grade levels (as in Exhibits 1-4) allows clearer comparisons with non- charter schools (which also vary somewhat in grade levels included).

One sidenote: Researchers working from the CDE research files (or any other resource that provides school-level summaries) are not limited to the mean-of-school-means analysis. Simple arithmetic indicates that multiplying the school mean by number of students provides the sum of all student scores in that school. And those sums can be combined to replicate grade 2-11 comparisons that start with individual scores. And because CDE also provides subgroup scores for numerically significant subgroups, the Exhibit 3 (SD in HighSD) comparisons can also be replicated and the Exhibit 2 (SD students) comparisons can be approximated (as over 90% of the SD students are members of numerically significant SD subgroups).

The justification for the analysis offered by Raymond is mostly correct.

> A more equitable way of comparing schools is to look at their ability to produce improvements in the academic performance of their students regardless of their nominal level of achievement. Gain scores as they

> are known assume that individual differences in student makeup and prior schooling will be reflected in the base score for an individual student, and any advancement will be due to the contributions of the school. Since there is no way to connect students with their individual prior histories, it is not possible to aggregate from individual student learning gains to arrive at a school-level score. Instead, differences in a school's API score over time serves as a crude proxy. (Raymond, p.16)

Describing progress is a good approach, and it is true that matched longitudinal data is not available. But Raymond is wrong in concluding that, as a consequence, the appropriate analyses are comparing means-of-school- means and analyses of covariance on school-means (see part B).

B. Misuses of Multiple Regression

Both Raymond and SKK attempt to augment their mean-of-school-means analyses with multiple regressions using a variety of predictors. (The SKK misuse of multiple regression was discussed in Rogosa, 2002, section 4). In particular, what Raymond terms an "Econometric Analysis" is in reality an assortment of very unthoughtful analyses of covariance for nonequivalent groups (charter vs non-charter) on school-mean data (see esp. Raymond, 2003, p.19 and App.C). This is an analysis that educational researchers have been warned away from for over 30 years especially in the Educational Evaluation literature. Raymond sells the multiple regression analysis by asserting:

> The second enhancement to the existing research comes from the use of multivariate analysis to relate differences in student background and school resources to both absolute performance levels and to school progress measures. This analytic approach allows us to study the effect of charter schools after removing potentially confounding influences that may differ across schools. (Raymond p.6)

> The estimated models present the opportunity to examine the way each of the factors affects performance holding all other factors constant. The models help put charter schools on an even footing with traditional schools by factoring out differences in the characteristics of students and their families and school resources. (Raymond p.20)

Not so. Regression adjustments for nonequivalent group comparisons (Raymond's "even footing") have been discredited in social science research for over 25 years (Cronbach et al. 1977; Weisberg, 1979). That is, with the California data it's possible and useful to describe academic progress for charter and non- charter schools. But it is not possible from these data to ascertain the counterfactual: What would have been the academic progress in charter and non-charter schools had the student populations and school characteristics been equivalent? The "as if by experiment" conclusion is not attainable by regression adjustments (Freedman 1991, 1997).

Moreover, interpretations of individual coefficients from a prediction equation such as the "Longitudinal Analysis" in Raymond (2003, App. C) should not be given serious consideration. The words of Mosteller and Tukey (1977, Chap 13, "Woes of Regression Coefficients") should be heeded:

> Regression is probably the most powerful technique we have for analyzing data. Correspondingly, it often seems to tell us more of what we want to know than our data possibly could provide. Such seemings are, of course, wholly misleading. (p.320)
>
> Part of the difficulty of giving meaning to regression coefficients arises because the coefficients themselves change depending upon which variables are present [ital] in the regression (p.300)
>
> Just dumping in a lot of closely correlated variables, and expecting a fit to the data to tell us, directly and simply, which one or ones are important usually expresses unjustified optimism. Any appearance this approach produces has a good chance of being misleading. It is far better to know what "these data cannot tell us" rather than erroneously to believe the results when they seem to have told us more than they actually can. (p.327)

Misuse of Multiple Regression Item 1: Residualized change scores. Not understood by Raymond is that the inclusion of "Last Year's API" as a predictor in the multiple regressions, Annual API Change Models (App. B) and Longitudinal API Change (App. C), effectively alters the outcome variable from amount of change to a residualized change score, a measure that researchers have been warned away from for over 20 years (Rogosa et al 1982, Rogosa & Willett, 1985). Raymond's justification for including the prior API is: "prior achievement is controlled (i.e. selection effect removed)" (p.22). Life would be nice if it were that easy. Residualized change scores address the question, How much would the school have changed if all schools had started out equal? rather than the answerable question, How much did the school improve?

Misuse of Multiple Regression Item 2: Regression Coefficient for Charter. Raymond's analyses of covariance (c.f. App. B and C) utilize a binary predictor for charter/non-charter, and the value of that coefficient is the analysis of covariance "treatment" effect (the standard adjusted mean difference). If charter and non-charters had been constituted by random assignment of students and resources to the two types of schools, then interpretation of this coefficient has justification. Sadly, in this highly non-equivalent groups comparison, Raymond mis-interprets this coefficient as indicating the relative efficacy of charters. For example, for the charter coefficient from the high school analysis with value 14.992, Raymond states: "Against all traditional high schools, however, charters showed significantly more positive gains--about 15 API points over a two-year span. This finding was significant at p>.05." (p.22). Note that Exhibits 1-3 consistently determine that improvement grades 9-11 is far less for charters that non-charters (margin of 14 points in Exhibit 2 and 21 points in Exhibit 3). Thus the analysis of covariance gets it exactly wrong.

Misuse of Multiple Regression Item 3: Demographic Variables

Raymond's results for the student demographic variables provide a striking example of the "econometric analysis" refuting the main feature of the API data. Raymond's Longitudinal API Change (App. C) regression uses as predictors both components of the SD (Socioeconomically Disadvantaged) classification, ParentEd not HS and Free/Reduced Lunch (NSLP), in a school-level form (see p.19). The elementary school regression yields the results below (excerpt from App. C), with both coefficients determined to be statistically significant:

Percent Free/Reduced Price Lunch	-0.378
Percent of Students with a Parent Not a HS Graduate	-0.970

Why are these results opposite to the main feature of the API data? Examine the grade 2-6 grouped results in Exhibits 1-3. Improvement increases notably in comparing improvement for all 2-6 students to improvement for SD 2-6 students and increases again comparing SD 2-6 students to those SD students in HighSD schools. A graphical comparison of the three sets of results is in Figure 2. Yet from Raymond's econometric analyses we are told that API improvement decreases as these variables, which measure "disadvantage", increase. And these effects are statistically significant. Come on. On a more general note, readers interested in useful analyses of relation between demographics and the API for all California schools are directed to the Interpretive Notes series (e.g., Rogosa 2000, 2001)

Misuse of Multiple Regression Item 4: School Size Coefficient

Based on the significant negative coefficients for School Size in the Longitudinal API Change (App. C) regressions, Raymond asserts strong conclusions:

> The conventional wisdom that size of school matters finds support in these findings. After controlling for other possible influences, the analysis reveals a significant association between smaller schools and higher achievement across all grade levels.(Raymond p. 6.)
>
> As shown in Appendix C, the results support the value of smaller schools. For each type of school, an inverse and significant relationship was found between size and progress: the smaller the school the greater the progress, all other factors being equal. (Raymond p.23)

As is often the case, the conventional wisdom is mistaken. Raymond's statements are much more a consequence of misinterpreting multiple regression, than reflecting any reality about California schools or student performance. In the California data there exists a strong confounding between size of school and student demographic variables that cannot be untangled by tossing terms into a multiple regression.

First consider the set of 93 charter schools. In the upper half of Figure 3 is plot of 2001 API versus number of students in 2001, and in the lower half, a plot of 1999 to 2001 API school improvement versus number of students in 1999. In each plot, School Type (E M H) is color- coded. In the top plot, small and medium-sized schools show the full range of API scores. The small set of very large schools (more than 1000 student scores in the API) show lower performance, with 3 very large schools having scores below 500, but two above 700. No striking trend is evident. In the bottom frame,

the main feature of the plot is that small and medium-sized schools show a large range of improvement and decline. The six middle and high schools with more than 1250 students in the API show little or no improvement, and those points will have extraordinary influence on any correlation or regression. As a counter-example consider the elementary schools of around 700 students in the API (twice the median sized of elementary school) showing improvement of almost 150 points. In sum, these charter school data do not provide evidence for a small school advantage, but the largest schools, which often face many demographic challenges besides just size, deserve more investigation.

Insert Figure 3

It may be helpful to recount some material from basic statistics courses in this context. The scatterplot from which Raymond bases her conclusions is not the bottom frame in Figure 3. The scatterplot the multiple regression uses is a plot that has on the vertical axis the residuals from predicting API improvement from all the variables in Appendix C except School Size using all schools (charters and non-charters). And on the horizontal axis in this plot are the residuals from predicting School Size from all the variables in Appendix C except School Size using all schools (charters and non- charters). See Neter et. al. (1996, Ch.9) and Mosteller and Tukey (1977, Ch. 13) for further explanation. Points in such a scatterplot defy interpretation.

Prior efforts to counter the misunderstandings about the relation between school size and API scores can be found in Rogosa (2000, 2001). For example, Rogosa (2001, Section 3, part A) displays the scatterplots of year 2000 API versus school size for each school type, concluding that

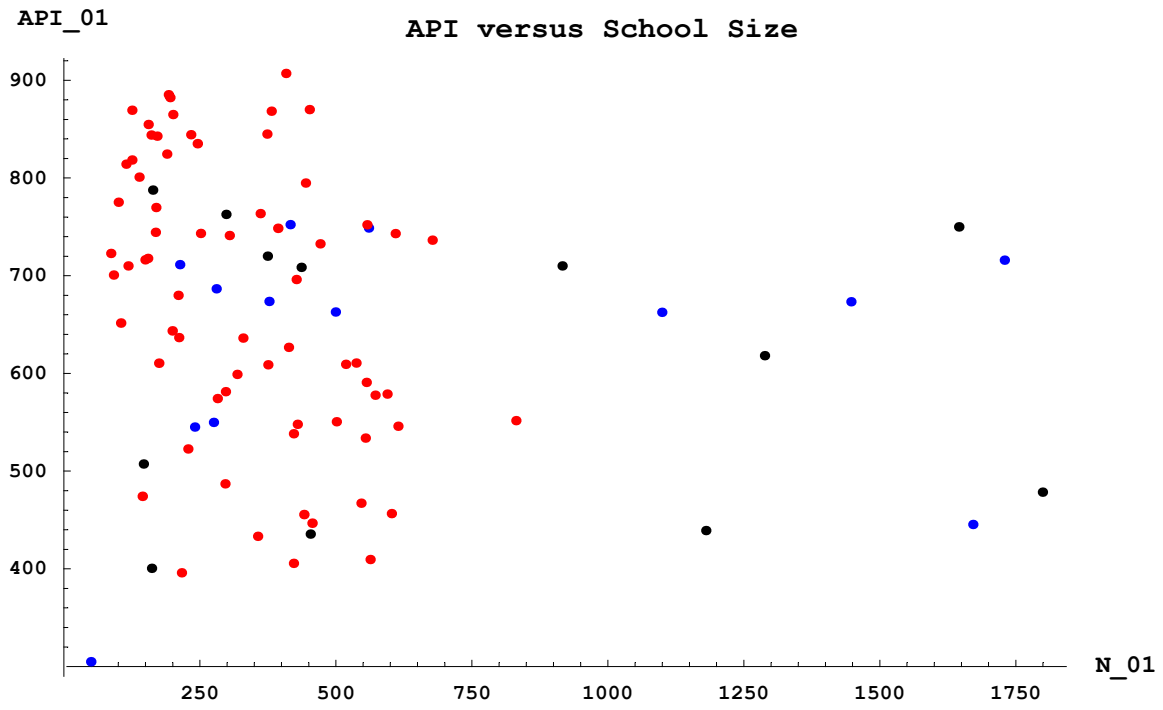
> From these plots one wouldn't conclude a notable relationship between school size and API. The High School plot reveals little dependence. For Elementary Schools, the ten schools with size above 1000 have strong influence on the correlation coefficient. This pattern is seen more vividly for Middle Schools where the six schools of size 2500 and above are very low scoring schools (API 450 and below) with the remaining 1100 schools of size below 2000 showing little dependence of API on school size. (Rogosa, 2001, p.32)

In addition, Rogosa (2001) Table 18 displays the distribution of school sizes at each statewide decile of API score and finds that school size appears rather uniform (in median and quartiles) across the statewide deciles. Except perhaps at the extremes, there is no apparent relation between API score and school size.

Moreover, there appears to be little relation between year-to-year improvement and school size. Figure 4 displays API improvement from 1999 to 2000 plotted against school size for High Schools and Elementary schools.

Insert Figure 4

Figure 3. School-size plots for the 93 charter schools; red(E), black(M), blue(H)



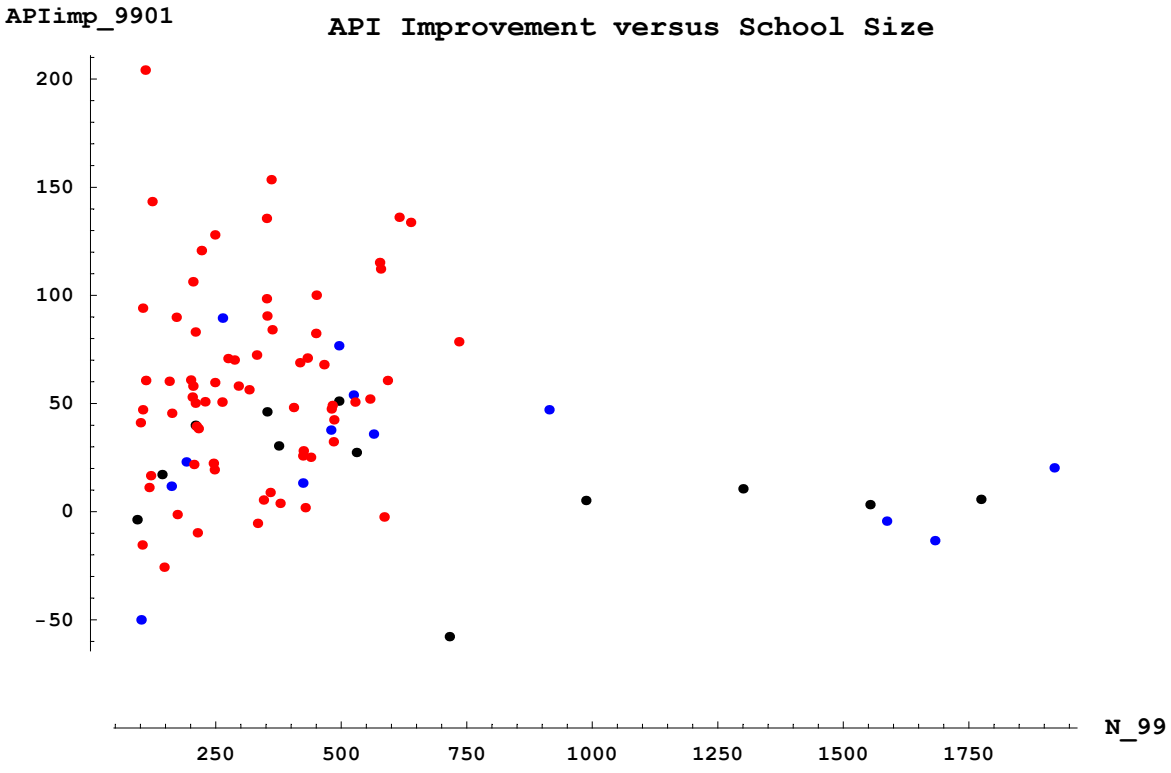
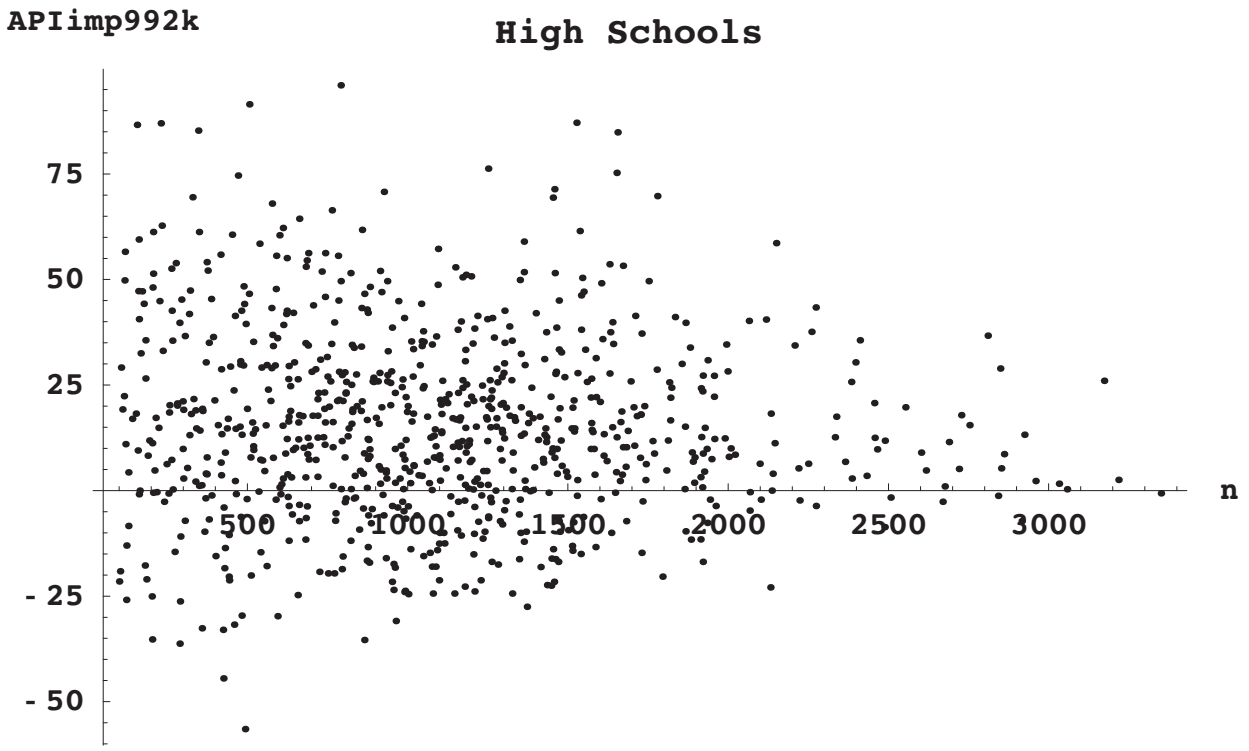
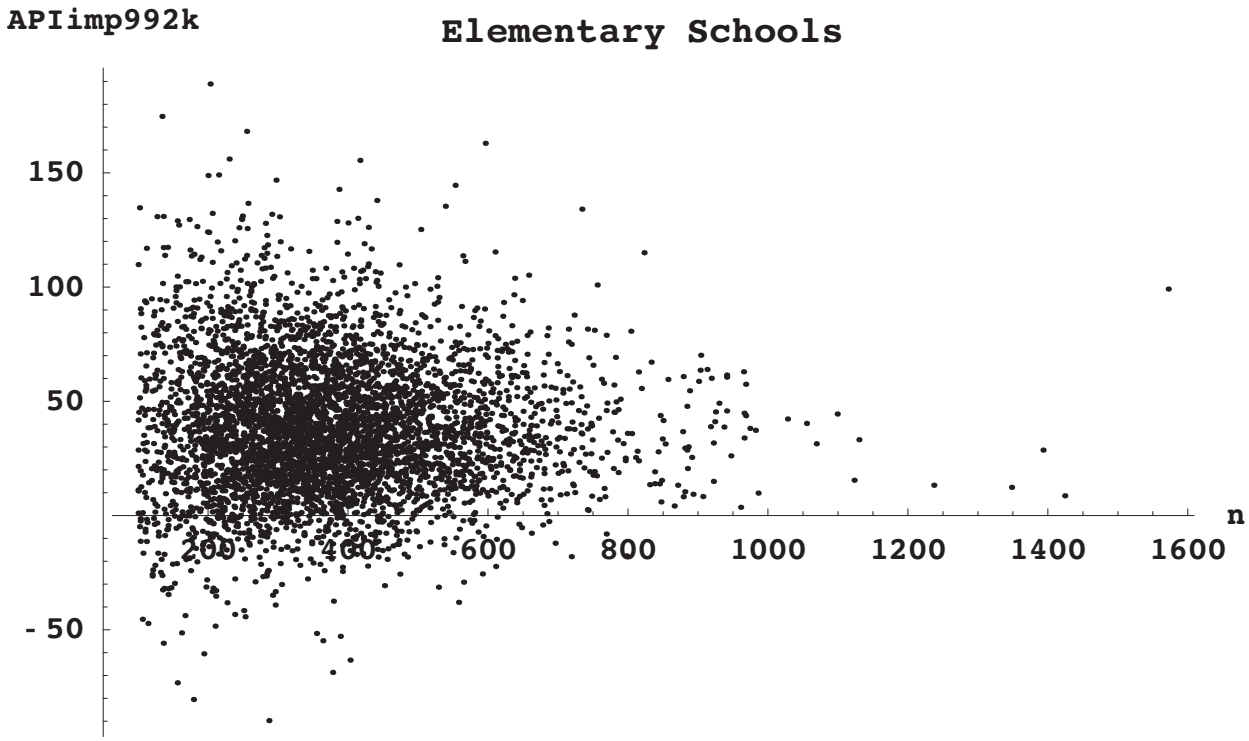


Figure 4. Plots of Improvement versus number of students





5. Conclusion.

The data analysis demonstrations in this report (like those in Rogosa, 2002), are mainly intended to guide researchers toward good approachers for these kinds of empirical research questions. The data analysis strategy of comparing progress in well-defined groups and subgroups, displayed in Exhibits 1-4 and Figures 1-2, indicates the overall conclusion (based on the test scores of successive cross-sections) that student academic progress in charter schools appears comparable to progress in traditional schools.

It's easy to agree with Raymond (2003) that:

> A more extensive analysis of the performance of charter schools is vital to help inform future decisions concerning chartering, accountability and choice in California public education. (p.5)

However, the empirical conclusions asserted by Raymond (2003) are not supported by the California API Data. Raymond's empirical assertions are similar to those of SKK who did the same mean-of-school means analysis for 1999-2001 over a year ago. As someone who personally is a strong supporter of charter schools, my real concern in doing both Rogosa (2002) and the present report has been that bad research on charter schools not seep into a perception about the quality of charter schools themselves.

References

Cronbach, L. J., Rogosa, D. R., Floden, R. E., & Price, G. G. Analysis of covariance in nonrandomized experiments: Parameters affecting bias. Occasional Paper, Stanford Evaluation Consortium, Stanford University, 1977.

Freedman, D. (1991). Statistical models and shoe leather. In P. Marsden, ed., Sociol. Methodology.

Freedman, D. (1997). From association to causation via regression. Adv. Appl. Math. 18 59—110.

Los Angeles Times, June 18 2003. "Charters' Test Gains Higher, Study Says", Erika Hayasaki, Times staff writer.

Mosteller, F. and Tukey, J. W. (1977). Data analysis and regression: A second course in statistics. Addison-Wesley.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models. Fourth edition. Homewood IL: Irwin, Inc.

Raymond, Margaret E. (2003). The Performance of California Charter Schools. CREDO, Hoover Institution, Stanford University May 2003. http://credo.stanford.edu/ Performance%20of%20California%20Charter%20School.FINAL.complete.pdf

Rogosa, D.R.(2000). Interpretive Notes for the Academic Performance Index California Department of Education, Policy and Evaluation Division November 2000. California Department of Education website: http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D. R. (2001). Year 2000 Update: Interpretive Notes for the Academic Performance Index. October 2001. California Department of Education website: http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D.R. (2002). A Further Examination of Student Progress in Charter Schools Using the California API. CRESST Technical Report 521, May 2002. available at: http://www.cresst.org/Reports/tr521.pdf

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50, 203-228.

Slovacek, Simeon P., Kunnan, Antony J. and Kim, Hae-Jin. (2002). California Charter Schools Serving Low-SES Students: An Analysis of the Academic Performance Index. Program Evaluation and Research Collaborative, Charter College of Education, California State University, Los Angeles. March 2002.

Weisberg, H. I. (1979). Statistical adjustments and uncontrolled studies. Psychological Bulletin, 86, 1149-1164.

Appendix A Listing of the 93 Charter Schools

CDS	SType99	SType01	SType02
01612596111660	M	M	M
10101086085112	E	E	E
10619946005730	E	S	E
10621661030642	H	S	S
10621666088942	E	E	E
10622406006704	E	E	E
10622406006712	M	E	E
10622406108328	M	M	M
16639336010466	E	E	E
16639906010557	E	E	E
16639906110233	M	M	M
19647256113146	M	M	M
19647331931070	H	M	M
19647331932128	H	H	H
19647331995836	H	H	H
19647336015929	E	E	E
19647336016323	E	E	E
19647336016570	E	E	E
19647336016810	E	E	E
19647336017016	E	E	E
19647336017040	E	E	E
19647336017057	E	E	E
19647336017156	E	E	E
19647336017677	E	E	E
19647336017701	E	E	E
19647336018063	E	E	E
19647336018204	E	E	E
19647336018634	E	E	E
19647336019145	E	E	E
19647336019236	E	E	E
19647336019525	E	E	E
19647336019715	E	E	E
19647336019806	E	E	E
19647336019905	E	E	E
19647336019939	E	E	E
19647336058267	M	M	M
19647336061394	M	M	M
19647336097927	E	E	E
19647336112536	E	E	E
19650946023527	E	E	E
21654176113229	E	E	E
24657716025654	E	E	E
28662666026900	E	E	E
28662666026934	E	E	E
28662666026983	E	E	E
29664156111413	E	E	E
29664156113138	E	E	E
30103063030632	H	E	E
30666216085328	M	M	M
30736356030183	E	E	E
31669443130226	E	S	E
31669513130168	H	H	H
31669513134657	H	H	H
31669516031363	E	S	E
31669516085252	E	E	E
33670586031959	E	E	E
34674056112643	M	M	M
34674396033799	E	E	E
34674476034508	E	E	E
34752833430659	H	H	H
36679343630761	H	H	H
36750773630837	H	E	E
37680236037956	E	E	E
37680236037980	E	E	E
37680236109771	E	E	E
37680236111322	E	E	E
37681063731023	H	H	H
37681303732732	H	H	H
37683386039457	E	E	E
37683386040018	E	E	E
37683386061956	M	M	M
37683386061964	M	M	M
37683386114961	E	E	E
38684783830411	H	H	H
38684786040935	E	E	E
41689996044333	M	E	E
41689996114953	E	E	E
41690056044473	E	E	E
41690216044721	E	E	E
41690216044739	E	E	E
41690216044754	E	E	E
41690216044788	E	E	E
41690216112213	E	E	E
42692786045918	E	E	E
44697996049720	E	E	E
44697996049829	E	E	E
48705734830113	E	H	H
49709536111678	E	E	E
49753586052369	E	E	E
50710926112965	E	E	E
50711006112627	E	E	E
50711346113286	E	E	E
50712176052922	E	E	E