The Achievement Gap in California: Context, Status, and Approaches for Improvement

Eva L. Baker, EdD - University of California, Los Angeles, Center for Research on Evaluation, Standards, and Student Testing (CRESST)
Noelle C. Griffin, PhD - UCLA, CRESST
Kilchan Choi, PhD - UCLA, CRESST

Introduction

This paper describes the achievement gap, its relevance to California education, and more specifically what might be done about it in the realm of standards, measures and relevant activities. We will illustrate what we believe to be the urgency of the California situation by working from the outside in, with the first tier of the argument displaying a comparison of the U.S. academic achievement performance to that of the international community. Second, we will look at achievement data within the U.S., specifically the results of the National Assessment of Educational Progress (NAEP), and describe the achievement of California students as measured by that indicator at different intervals. We will examine comparative status and rates of change in reading and mathematics overall and for subgroups of students. Finally, we will concentrate our attention on California itself. Within California we will consider outcomes and subgroup performance across a range of indicators. While most of our emphasis will be on the tested grades in elementary and secondary schools, we also plan to explore illustrative data from higher education and preschool levels. Although the preponderance of the data we will cite comes from descriptive and archival sources related to current accountability requirements, we will include studies that have examined approaches that may have implications for improvement and for additional measures of effectiveness.

The charge to the authors of this document was to understand the gap in the light of the measures and assessment policies now in place, which inevitably brings in a consideration of standards. To place California in context once again, we will present a top-level history of how federal policy has addressed lower-performing students. We will describe the complexities, both positive and negative, of different approaches to measuring proficiency accurately for each and every student. Taken together, these concerns address the validity, accuracy, fairness, and interpretability of findings in the light of the existing state standards, the range of purposes intended for these measures, and the confidence that can be placed in the obtained results. We conclude that the measurement system in California needs revision. This change is not primarily driven by the system's technical aspects, although there is a lack of evidence

This is one in a series of ten papers and policy briefs prepared through a collaboration between the California Department of Education and the University of California organized by the UC Davis School of Education Center for Applied Policy in Education. April, 2008.

supporting the efficacy of its information for key purposes. We will end with the policy recommendations addressing standards, measures and interpretations that may serve as more powerful to interventions intended to reduce the gap. What should be clear in the ensuing discussion is that the differences found among subgroups are not artifacts of measurement nor can be ethically eliminated by simply changing measures, indicators, and cut scores. However, if measurement is a cue to practice, there may be important ways to create greater intellectual and social coherence in schools and to provide the everyday supports teachers need so that all students, against many odds, may attain the knowledge, skills, propensities, and understandings that will help them all succeed in school and beyond.

What Do We Mean by Measured Achievement Gap

When two or more groups have severe, persistent discrepancies in formally measured academic performance, the phenomenon is commonly called an "achievement gap." A one-time, moderate difference may not qualify as a "gap," for the term typically connotes that for an identifiable subgroup performance differences are obvious and relatively severe on common measures and indicators. A second criterion, gap persistence, suggests that the discrepancies among performance of compared groups are empirically intransigent, that is, they have not been overcome by available interventions. A possible misconception, or worse, a pernicious interpretation, of the general term achievement gap involves the belief that performance for each group clusters around its own average and does not overlap with the scores of higher performing groups. Were that to be the case, the lack of overlap might spur beliefs of limitations; to wit, that individual members of lower-performing groups would not reach regions of achievement occupied by the high performers. The reality is that there is considerable overlap on a wide variety of measures that documents that members of lower performing subgroups achieve in the higher ranges. The overlap also goes the other way, with plenty of evidence that many low performing individuals come from higher performing groups. In other words, overlap among groups is a fact and it is a mistake to reckon that membership in a subgroup implies that individuals share strengths and weaknesses.

It is for this reason that it is important to think of students as individual learners, focusing on the operational evidence of their performance, and adopting the perspective that each student will be able to meet desired standards. The persistence of gaps suggests that individual progress is not necessarily easy, but will depend upon the acquisition of prerequisites, motivation, and cognitive readiness [1] to meet academic challenges and rapidly changing expectations in service economies. We will return to this topic later.

In addition to gap definitions and their implications for the beliefs of students, teachers, and the public, we should consider gap reduction in a much more refined way than "the gap has opened" or "the gap is closing." It may be that the "head-start" of higher performing groups is not easily overcome. To close the gap in common parlance would mean that the rate of performance of the lower achieving groups would need to greatly exceed the higher performing group. There are statistical reasons (regression toward the mean) that support the idea of somewhat faster performance of lower achieving students

(that is, they are presumably encountering "easier" material, but perhaps not experientially) and the slower progress of higher performing students (as they try to master very difficult, complex material). However, it is fair to say that this statistical argument holds only when the conditions in which learning occurs are equivalent, or at least, each student independent of group membership could randomly experience a range from excellent to poor learning environments. Our inspection of process and resource data, to be dealt with in depth by other writers (and discussed broadly later in this paper), suggests that conditions are generally demonstrably different for lower and higher performing students. There is evidence that students in lower performing groups show different progress in schools that have higher performing students in the plurality. We also see greater variation within schools than between schools.

So taking a benchmark approach to closing the gap would mean that, first, divergent rates of progress would need to be equivalent (imagine students progressing on higher or lower but parallel lines). If lower achieving students increased their rate of improvement, then these lines could conceivably meet or cross. We think it might be wise to consider the phases of progress that would need to be marked to indicate that substantial, sustained progress is being made. It may be that "gap" closing is in reality too big a jump to consider in a three-year period, for instance.

However, it might be that gap closing around a particular, targeted important set of outcomes would be feasible. For such identifiable skillsets or understandings, we might expect more rapidly, converging lines of student group performance. One milestone might be looking at the growth in the overlap of lower performing groups. This growth needs to be inspected not only in the frequency of students who move between categories (e.g., Basic), but also in the actual performance on subscales of the target examinations. In that case, we could tell if progress was being made that did not yet show up in approaches based on overall student achievement classification. There are critics who argue that addressing the lower group's performance always comes at the expense of increasing the higher group's performance. As a consequence, parents may shop for better schools in either the private or public sectors. Is there a way to achieve our goals and balance the interests of all children and their parents? These questions, although bald and uncomfortable, need to be confronted and resolved.

The History of Policy to Reduce the Gaps Among Group Performance

Two contrary views may be simultaneously held when considering such gaps in achievement. One is that persistent and resistant gaps are caused *principally* by out-of-school factors, such as parental background and economic status [2,3]. The other holds that differences in performance can be substantially reduced by management action, in other words through dicta of educational policy (No Child Left Behind [NCLB] [4]). If one buys into the first view, of the incontrovertible structural sources impinging on performance, at least two inferences can be made: that early background plays an essential and invariant role (see meaningful differences) and that "catching up" is an unlikely outcome without fundamental social change. Lack of complex language, parental reading, and other behaviors often associated with poverty forecast expectations for individual members of under-performing groups, implying differential and more limited potential and growth for these students. On the other hand, one can decide that the correlation between background and performance must be vitiated, if not

eradicated, by an emphasis on intervention. In California, preschool initiatives provide early attention to learning for some [5,6]. Adult mentoring has been suggested as an essential part of raising and supporting higher aspirations of students and building resiliency [7].

In the educational system, the second view, that policy can change performance, has operated for more than 40 years. Policy-initiated interventions by the federal and state governments have over the decades at one time embraced special tests and separate instructional treatments for lower performing groups. They emphasized the acquisition of minimum competencies, as described in the various guidelines by the reenactments of the Elementary and Secondary Education Acts in their first two decades of operation. Since the late 1990s, the policy focus has shifted to high performance, expressed rhetorically as challenging standards and assessments for all students, and for some, as college education for all. This high standards argument was engendered by the recognition of the emerging global economy, the lackluster placement of American students in international comparisons (one reason for a subsequent section of this paper), and the importance of attaining world class educational standards.

Political and policy momentum for standards and assessments came from a series of high-profile enterprises, generated by the Charlottesville meeting of the Governors in 1989; the National Educational Goals Panel [8], the National Council on Education Standards and Testing [9] and its report *Raising Standards for American Schools* conducted under the administration of President Bush the elder. The subsequent Elementary and Secondary Education Act of 1965 (ESEA) reauthorization, *Improving American Schools Act* (IASA) of 1994 [10] emerged in the Clinton administration. It was acknowledged that its emphasis on standards and assessment (at grade levels more or less compatible with NAEP administrations) was federal legislation that substantially reified reform efforts of leading "innovative states," a subset that included in the late 1980s and early 1990s: California, Connecticut, Kentucky, Maryland, and by some sights, New York, and Texas. In these states, the investment was in goals (subsequently renamed standards), carefully developed curriculum frameworks, and challenging assessments. These were exemplified in California in the California Learning Assessment System (CLAS), in Kentucky Instructional Results Information System (KIRIS), and now the Massachusetts Comprehensive Assessment System (MCAS), all with a strong emphasis on complex learning in subject matter. The enactment of NCLB in 2002 gave us IASA on steroids, by adding grade levels for testing, annual targets, sanctions for inadequate performance, and unambiguous language (e.g., "failing schools"). NCLB, the grandchild of the reforms in the early 1990s, also added strong expectations for the testing of English Language Development and a state-by-state plan for target setting and attainment, Adequate Yearly Progress, that had laudable goals at its heart: making sure schools address the participation and performance of all subgroups by disaggregating from the overall performance, and presenting results by identifiable subgroups in states' annual reports. Unfortunately, problems arose in its mechanics. It was predicted and then shown that schools might be labeled "failing" for a number of reasons, for instance, because they had too many subgroups, or had large proportions of transient students and teachers (raising the question of how much a "schools" performance in 2002 had to do with the same school in 2006 with its

substantially different players: administrators, teachers, and students).

California had earlier adopted an accountability approach, the Academic Performance Index [11, 12], a school level measure that weighted California state examinations, identified targets, and provided comparisons in status and progress to sets of schools thought to be comparable in demographic, size and other factors. Another consequence of the NCLB emphasis on census testing had been anticipated in California by Governor Pete Wilson's insistence on individual scores for each student rather than sampling a broader range of content with an approach that would not yield comparable scores for students. The result of this decision in part was the reappearance of multiple-choice examinations as the principal mechanism to measure performance. This approach is relatively inexpensive compared to then current methods of scoring student constructed, longer and more complex responses. It also supported the strategy taken in California to have relatively many state standards and to lightly sample performance across a number of them, rather than going in depth in a few areas. Because of leadership changes in the California Department of Education and strong criticism of CLAS (because of content, privacy, and technical questions), the early "innovative" state status of California was rapidly submerged. Multiple-choice tests also provide the possibility of making sure that growth expectations were about the same at different grade levels (vertical scaling).

Consequent emphasis on standards and many more sanctioned tests put attention on teachers' ability to use data driven analysis in their instructional planning. Without chronicling the many efforts that California has made in professional development, curriculum adoptions, incentives for high test scores, districts' mandated use of particular materials and the like, it is clear that California has been active and concerned about the quality of learning of its students. What have been the effects on student learning?

Is There a Serious Achievement Gap?

Sadly, the California activity has not resulted in good news. To the question "Is there a serious achievement gap?" the answer is yes. One can observe gaps between higher performing groups and lower performing student groups in both status and rates of change using any one of a range of indicators. To give some depth to the meaning of these differences, we will contextualize them briefly with data from other settings. The point of this review is that if we want California students to be economically competitive in a global and in our national economy, we need to understand where they now stand and need to go. Working from the outside in, we will first illustrate the place of U.S. performance on international comparative examinations, then California's standing in national measures in the U.S., and finally address achievement differences within California.

The International Context

Although the research in the area of international comparisons is extensive and not without its critics (see [13,14,15]) we begin our brief discussion of gaps by looking at the performance of the U.S. in but one of many studies, the Programme for International Student Assessment (PISA). PISA comparisons focused on the mathematics performance of 15-year-olds in

more than 40 countries, 30 of which were members of the Organisation for Economic Co-Operation and Development [16] (see [17]). Collaborative content for the examination was developed through an iterative framework to guide test items and task selection. Although the focus of the examination was mathematics in 2003, one of its important components was cross curricular definitions of problem solving skill, intended to transfer and apply to other content areas, one of the definitions of cognitive readiness. While we will not report the full findings, which includes copious information about settings, they can be easily accessed on the OECD website (http://www.oecd.org/).

How does the U.S. perform? In a nutshell, the U.S. ranked 25th among the 30 OECD participating countries in 2006, slightly lower than its 23 of 29 ranking in 2003. In 2006, U.S. students scored 474 points on a scale with an OECD average of 498. At the top of the distribution were Taipei, Finland, Hong Kong-China, Korea, the Netherlands, Switzerland, and Canada. At the lowest end of performance were Kyrgyzstan, Qatar, Tunisia, Brazil, Colombia, Argentina, and Jordan. Not statistically different from scores of U.S. students are Croatia, Portugal, Spain, Azerbajian, and the Russian Federation. A larger percentage of U.S. students (54.2%) scored below the third level of 7 proficiency levels than the OECD average (43.2%) and only a little less that 8 percent reach level 5 or above, compared with the OECD average of 13 percent.

When 2006 performance is adjusted for levels of Gross National Product, school expenditures, and responses to the economic, social and cultural status (ESCS) scale, U.S. students' performance predictably drops, showing that they are underperforming their peers in countries with similar economic characteristics (see Table 1). Comparisons were also made on an "equity" dimension, where the effects (of different economic status) within countries were compared. The metrics used compared the difference in mathematics score points associated with a .5 standard deviation change on the ESCS measure. The impact was computed by school ESCS average and for individual students. The U.S. ranks fourth in ESCS score. For the U.S., school intake characteristics accounts for about 30 score points for each .5 standard deviation on the ESCS measure (comparable to the OECD average), whereas individual status accounts for about 15 points on the mathematics examinations (high compared to OECD countries). In summary, using the data above, the U.S. performance is below the average rank of all countries and considerably below OECD member countries. The U.S. underperforms for various measures of economic well-being, and shows higher impact of individual student socioeconomic status. Explanations of U.S. performance levels include the size and diversity of our population, our diversity and continuing broad base of immigration, and performance-based school admission processes in other countries, among other reasons.

PISA 2006 also reported findings in science of students within a given country related to their immigrant status, the closest variable to ethnicity incorporated into PISA analyses. The overall OECD average in science was 506, and the average decrease in performance attributed to first-generation students was 58 points. For students who are second generation, born in one country with a parent born in another country, the average score difference was 55 points lower. Comparable numbers for the U.S. were 57 for first generation and a 43 point lower score for second-generation students when compared with native-born students. While these numbers are somewhat better than the OECD average, they still represent a large,

Table 1

Economic and social indicators and the relationship with performance in mathematics

Countries	Mean performance on the mathematics scale	Economic and social indicators				Adjusted performance on the mathematics scale			
		GDP per capita *(In equivalent US dollars using purchasing power parities)*	**Percentage of the population in the age group 35-44 years that has attained at least upper secondary education**	**Mean PISA index of economic social and cultural status (ESCS)**	**Cumulative expenditure per student between 6 and 15 years** *(In equivalent US dollars using purchasing power parities)*	**Mathematics performance adjusted by GDP per capita**	**Mathematics performance adjusted by GDP per capita and educational attainment**	**Mathematics performance adjusted by the mean PISA index of economic social and cultural status**	**Mathematics performance adjusted by cumulative expenditure per student between 6 and 15 years**
Australia	524	26 685	62	0.23	58 480	516	528	509	520
Austria	506	28 372	82	0.06	77 255	493	487	501	489
Belgium	529	27 096	66	0.15	63 571	520	529	519	522
Canada	532	29 290	86	0.45	59 810	518	510	502	528
Czech Republic	516	14 861	91	0.16	26 000	536	504	505	534
Denmark	514	29 223	81	0.20	72 934	500	496	501	501
Finland	544	26 344	85	0.25	54 373	537	525	528	543
France	511	26 818	68	-0.08	62 731	502	508	516	504
Germany	503	25 453	86	0.16	49 145	498	484	492	505
Greece	445	17 020	58	-0.15	32 990	460	463	455	458
Hungary	490	13 043	79	-0.07	25 631	514	492	495	508
Iceland	515	28 968	62	0.69	65 977	501	517	469	506
Ireland	503	29 821	65	-0.08	41 845	487	500	508	510
Italy	466	25 377	50	-0.11	75 693	460	483	473	450
Japan	534	26 636	94	-0.08	60 004	526	506	539	529
Korea	542	15 916	79	-0.10	41 802	560	541	549	549
Luxembourg	493	w	w	w	w	w	w	w	w
Mexico	385	9 148	26	-1.13	15 312	419	444	461	410
Netherlands	538	28 711	71	0.10	55 416	525	531	531	536
New Zealand	523	21 230	80	0.21	m	528	515	509	m

Norway	495	36 587	91	0.61	74 040	463	459	454	481
Poland	490	10 360	48	-0.20	23 387	521	526	504	510
Portugal	466	17 912	20	-0.63	48 811	479	521	508	468
Slovak Republic	498	11 323	91	-0.08	14 874	527	490	504	523
Spain	485	21 347	46	-0.30	46 774	490	511	505	489
Sweden	509	26 902	87	0.25	60 130	500	487	492	504
Switzerland	527	30 036	85	-0.06	79 691	510	504	530	508
Turkey	423	6 046	25	-0.98	m	465	487	489	m
United States	**483**	**35 179**	**88**	**0.30**	**79 716**	**454**	**451**	**463**	**465**

Source: http://www.oecd.org/dataoecd/1/60/34002216.pdf

although diminishing, difference. For comparison's sake, in Australia, with a small population but a high level of immigration, the difference between first-generation and native born performance was 3 points, and 2 points between native and second-generation students. In Canada, the difference between native and first-generation students was 22 points, and between native and second-generation 12 points. In the United Kingdom the respective decreases were 41 and 26 points for first- and second-generation students, compared with native born students. In conclusion, we point to the obvious. The U.S. is not in a leadership position on the PISA measure.

The National Context

We now turn to national findings regarding the achievement gap. The measure used nationally to monitor student achievement since the early 1990s is The National Assessment of Educational Progress (NAEP) [18]. NAEP assessments are implemented across multiple grades and subjects, but as part of the Nation's Report Card overall and subgroup findings are reported for grades 4 and 8 in math and reading (http://nces.ed.gov/nationsreportcard). For key student subgroups, we will address both overall trends in NAEP achievement gaps and how results from California compare to other states. It is also worth noting that California is in the lower ranks internationally. For example, in a recent analysis that equated international results from the Trends in International Mathematics and Science Study (TIMSS) with NAEP scores, California scored at approximately 1/3 of the top country's score [19].

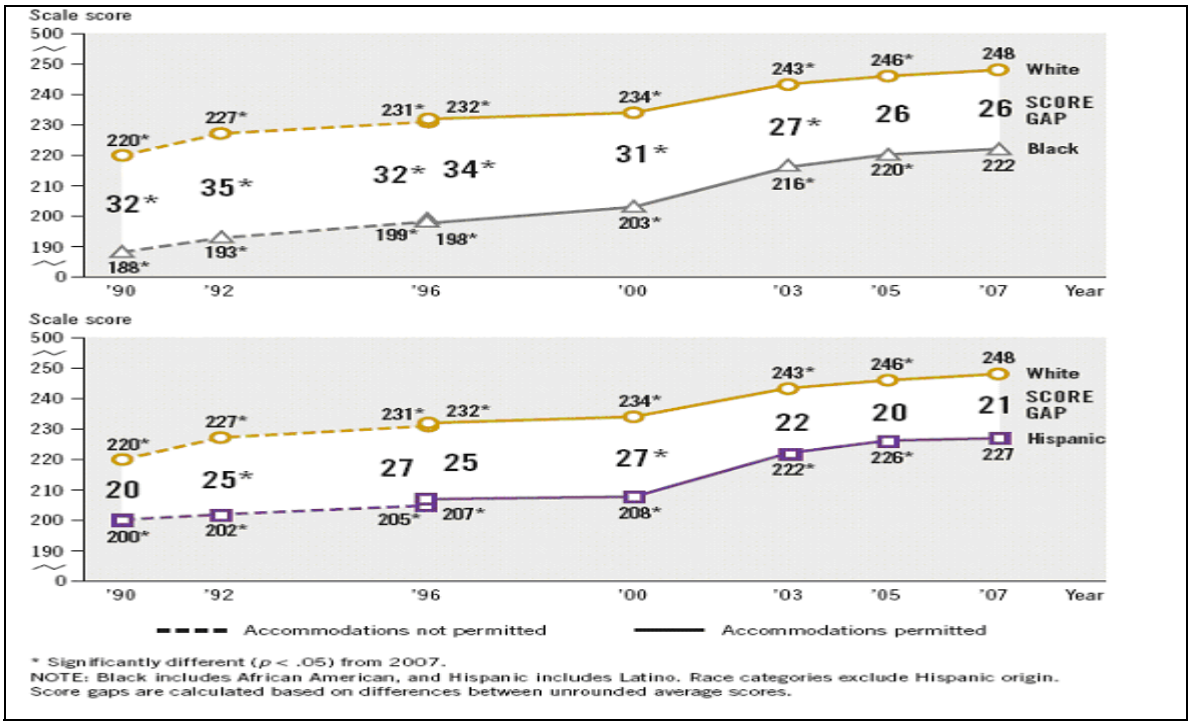
California Performance

California's performance has been in the lower ranks of States overall for the years we examined, 1990, 1998, 2002, 2003 and 2007, for example 48th ranked in grade 4 mathematics, 46th ranked in grade 8 mathematics, and 49th in both grades 4 and 8 reading in 2007.

African American Students

Gaps in achievement between African American and white students have existed since the beginning of NAEP for both math and reading, although both groups have shown increases in NAEP scores across this timeline. The size of the gap has also varied over time. For example, Figure 1 shows the trend in grade 4 NAEP mathematics average score and score gaps by selected racial/ethnic groups including white and African American. The math score gap between white and African American students decreased overall by 6 points from 32 (1990) to 26 (2007), although there was no change between 2003 and 2007. Similar gaps persist at grade 8, with average score gaps between white and African American students in math at grade 8 much larger than the average grade 4 score gap. As with the grade 4 students, there is some decrease in math score gap between grade 8 white and African American students across the years, for example, a drop from a 40 point gap in 2000 to a 32 point gap in 2007. However, the gap between white and African American students remains large, and, although the general trend is in the right direction, the decreases are usually well within the chance range.

NAEP reading results for African American students show some similar trends. As an example, Figure 2 shows the grade 4 NAEP average reading score and score gaps, by selected racial/ethnic groups. While overall scores for both African American and white students have increased modestly across the testing period, the reading score gap between the groups has decreased 5 points. At grade 8 reading scores have also

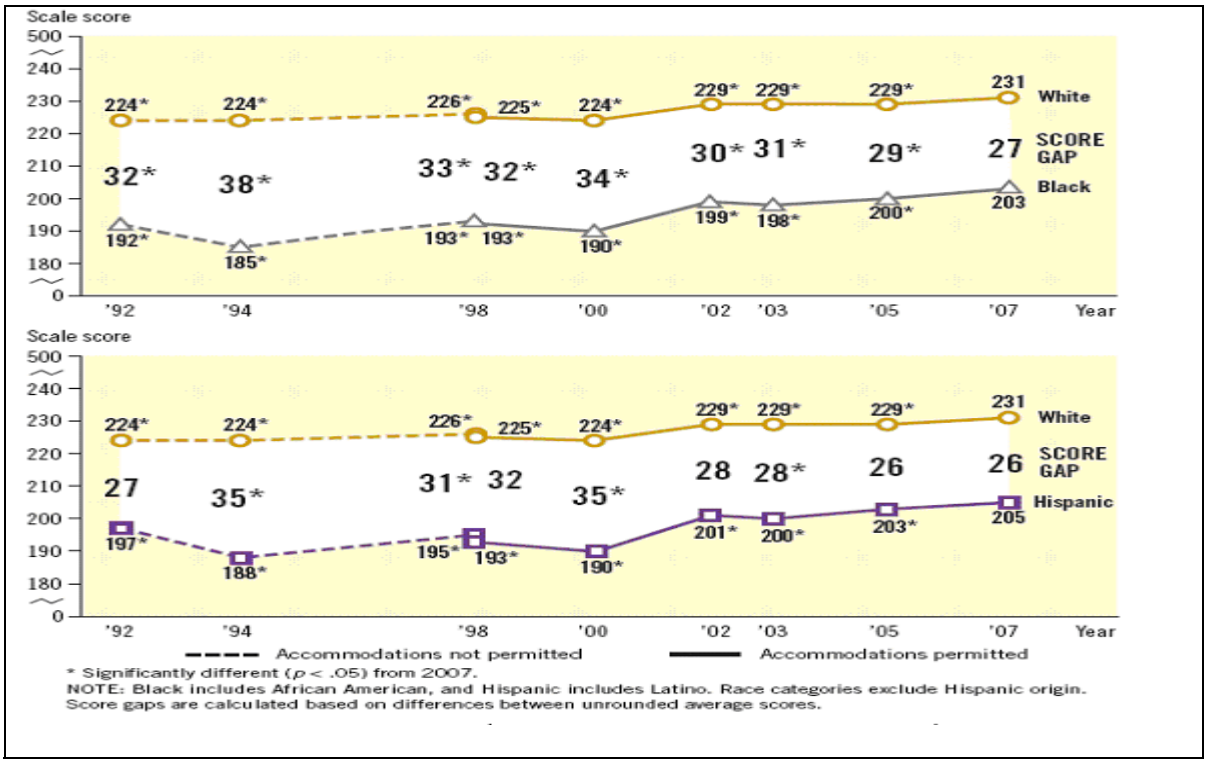


Source: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007494

Figure 1. Trend in grade 4 NAEP mathematics average score and score gaps, by selected racial/ethnic groups.

increased slightly over time for both groups (white students have gone from a score of 267 in 1992 to 272 in 2007, and African American students from 237 in 1992 to 245 in 2007). However, reading score gaps between grade 8 white and African American students have not significantly decreased since 1998, staying in the 26-28 range across that time period.

California's overall pattern of under-performing compared to other states was consistent across math and reading for African American students in 2007 with one exception—as shown in Figure 3, grade 4 African American students in California did not score significantly different than the national average. Furthermore, a handful of other ethnically diverse states show significant progress in reducing gaps between African American students and white students. For instance, in reading at grade 4, states with relatively high proportions of African American students have demonstrated larger reduction in differences in reading between white and African American grade 4 students (in approximately 10 scale points, New Jersey and Michigan from 2004-2007). In grade 4 math, relatively large decreases are also found for diverse states such as District of Columbia, Rhode Island, Michigan, Pennsylvania, and New York. At grade 8 level, where growth typically slows, Colorado posted a 13 score point gap decrease between 2003 and 2007 on grade 8 math, and Kansas and Florida decreased differences by 10 and 8 points respectively. In other terms, at this rate Colorado could overcome differences in performance in two NAEP cycles (2 years each) while it might take Kansas and Florida three cycles to do so. (Given the present NAEP rate of change, California's gap closing for African Americans and whites would take closer to



Source: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007496

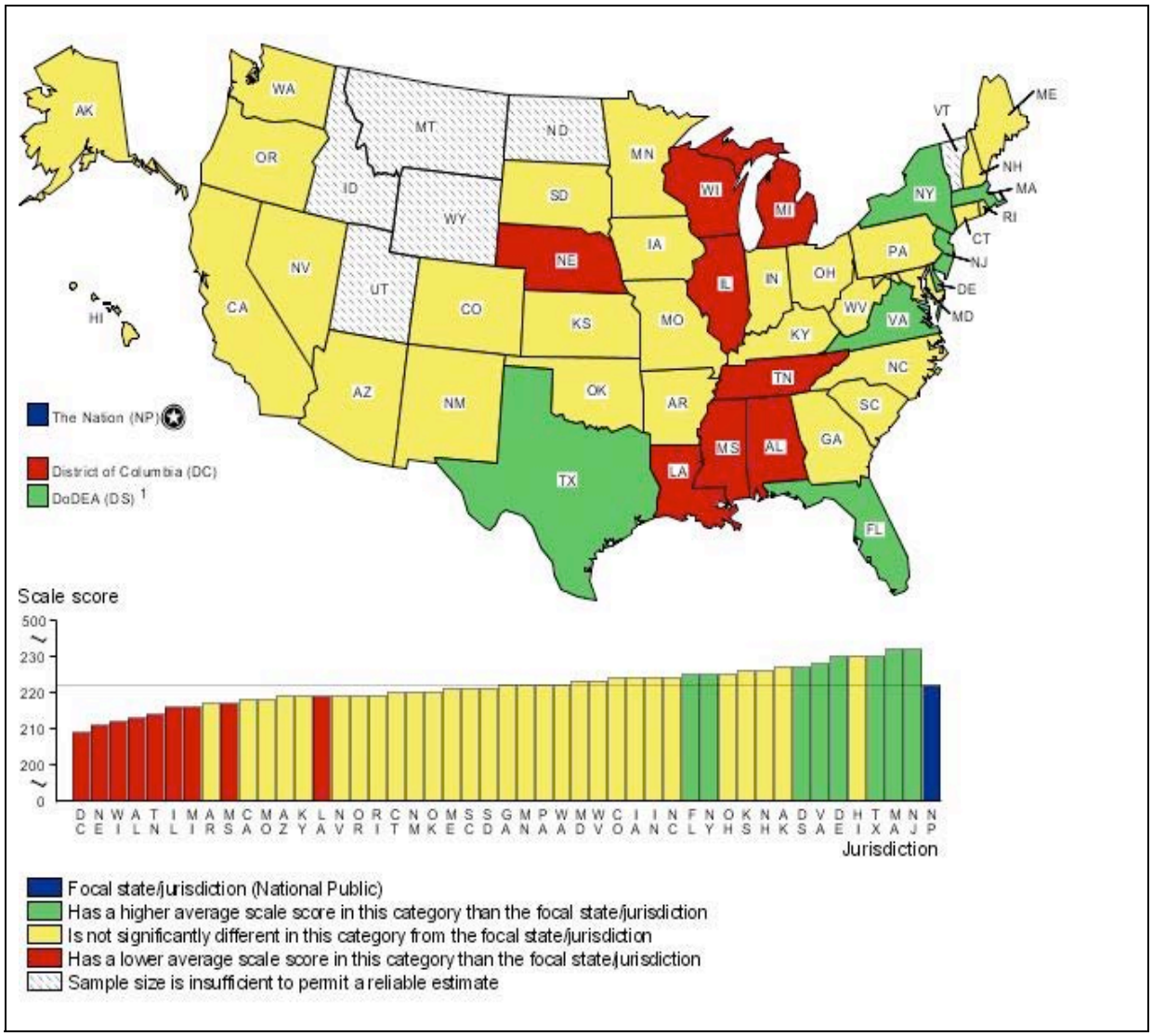
Figure 2. Trend in grade 4 NAEP reading average score and score gaps, by selected racial/ethnic groups.

10 times the length of time projected for the higher performing states.) Note also, for example, that Colorado scale scores for both whites and African Americans exceed those of California, so there progress is not coming at the expense of higher performing groups.

Hispanic Students

As with African American students, while overall NAEP scores for Hispanic students have increased over time, the gaps between white and Hispanic students persists. As Figure 1 shows, grade 4 NAEP math scores have increased across the testing period for Hispanic students (from 200 in 1990 to 227 in 2007), but the score gap between white and Hispanic students has changed at several points throughout the years. The gap slightly increased and then decreased again starting in 2000. However, there has been no statistically significant change in gap between white and Hispanic for the last four years, from 2003 to 2007. As with grade 4 students, there are gaps evident throughout the NAEP timeline between white and Hispanic students for grade 8 math, with no statistically significant change in gap between white and Hispanic students between 2005 and 2007 [20].

Again, as with reading, overall scores for Hispanic students have increased across the years, but the score gap between white and Hispanic students persists. For example, for grade 4 reading the gap only decreased 1 point overall between 1992 and 2007, and although it has fluctuated in that time, the change between 2005 and 2007 is attributable to chance (i.e., the margin of error in measurement). There has been even less variation in the nature of the Hispanic-white gap at grade 8 reading, which has



Source: http://nationsreportcard.gov/math_2007/m0005.asp?subtab_id=Tab_5&tab_id=tab1#chart

Figure 3. Cross-state comparisons of average grade 4 NAEP mathematics scores in 2007 for African American students.

ranged 24-27 points across the years and currently is 25. As noted above, Hispanic students follow a similar pattern to that described for the state overall in comparison to the national average, with the Hispanic students in the state performing below the national average in both math and reading. Once again, there are ethnically diverse states that show more dramatic reduction in scale score differences since NAEP began on math and reading for Hispanic students. For example, in grade 4 reading, the District of Columbia had a gap decrease of 15 points, Delaware 9 points, and Michigan and Illinois 7. As important perhaps as decreases in differences are the overall levels of performance of Hispanics in comparison states. Hispanic students in Florida exceed California student performance in grade 4 reading by 23 points, Texas students, by 17 points, New Mexico students by about 9 points, and Colorado students by 14 points.

At grade 8, the Hispanic students in the Defense Overseas Dependent Schools (DoDEA) achieve more about 34 score points above California Hispanic students, with only a 5 scale sore point difference between white and Hispanic students. Ohio Hispanics scores on grade 8 reading exceed those of Californian Hispanics by about 20 scale score points, and Texas and Colorado by more than 10 points.

Comparisons between African American and Hispanic students add some additional information to the picture. For example, in grade 4 mathematics there is no difference on NAEP scores found between the two groups, but, by grade 8 Hispanics outperform African Americans by 3 points. In reading grade 4 African American students have a 5 point superiority, but the balance switches by grade 8 to a 2 point advantage by Hispanics. One could conclude from the data that there is a relative acceleration of performance for Hispanics when compared with African Americans in both reading and math.

English Learners (ELs)

An achievement gap is also evident in NAEP results when comparing ELs to other students. Overall both grade 4 and grade 8 public school EL students have substantially lower scores in both NAEP math and reading that their non-EL counterparts (see [21]). For grade 4 students, score gaps in math and reading between EL and non-EL are respectively 30 and 35. For grade 8 students, score gaps in math and reading between EL and non-EL students are respectively 37 and 41.

Other Subgroups

Other group differences in NAEP scores are of note. First, achievement gaps are presently based on student socioeconomic status (SES), with the score of students not eligible for free/reduced lunch higher than those who are eligible, and students who are eligible for reduced lunch scoring higher than those eligible for free lunch [22]. In these comparisons, California again does poorly, the 46th ranked state in grade 8 reading and mathematics, ranked 50th in grade 4 math and 49th in grade 4 reading. Differences in performance by gender have been reported. At both grades 4 and 8, girls scored higher than boys on reading, with a slight opening of this gap. In mathematics, boys 2 point at grade 4 advantage drops to zero in grade 8.

Summary

The national picture is fairly consistent overall across grades and content areas (math and reading) in terms of gaps in achievement, but varies substantially within states. Although scores for all students groups have increased over time, gaps between white students and both African American and Hispanic students have generally persisted, particularly when looking at grade 8 students and considering the positive relationship between minority status and poverty indicators [23]. Similar trends exist when comparing ELs to their non-EL counterparts. California on the whole performs lower than the national average for the majority of student subgroups on both math and reading, and several ethnically diverse states further out perform California in gap reduction.

Findings from California

The purpose of this section is to examine trends in achievement gaps evident in existing California data. The goal is not to provide an exhaustive detailing of all subgroup differences within all measures, but rather to identify and highlight the most noteworthy findings and trends. Several measures will be discussed. The API is described earlier. The California Standards

Tests (CST) are given in a range of content areas; the discussion below focuses on CSTs in English Language Arts and mathematics, which are completed by grades 2-11 and 2-9, respectively [12]. The California High School Exit Exam (CAHSEE) in math and language arts is offered to students starting in grade 10. Beginning with the class of 2006, passing has become a graduation requirement [24]. Each of these measures is discussed below in relation to specific student subgroups.

African American Students

Tables 2 and 3 show the 2006 base and 2007 growth API scores by grade clusters and subgroupings. African American students score lower than all other ethnic groups across all grade levels across both years, although the gaps were slightly smaller in 2007 than 2006. There is some fluctuation in the gap across grade levels, but no linear increase. For example, in 2007 there was a 169 point gap between white and African American groups at grades 9-11, and a 178 point one at grades 7-8. Although this shift suggests some consistency in the achievement gap, possible changes in the student population over time need to be considered when interpreting these results. For example, dropout rates (grade 9 and above) tend to be higher in student groups on the lower end of the achievement gap, such as Hispanic students and African American students [25]. Considering that students who drop out tend to be on the lower end of the achievement spectrum [26], it can be inferred that the African American and Hispanic student subgroups at the upper grades have disproportionately lost their lower achieving students compared to other groups. That is, if anything the gaps at the highest grades should be substantially reduced by differential dropout rates.

Looking beyond the API at the CSTs, Tables 4 and 5 display the overall (i.e., across grade levels) percent of students scoring proficient or above by subgroup for both CST math and language arts from 2003-2006 [12]. Although there is an increase in percent proficient for all groups from year to year, similar gaps to the API are seen in both content areas. That is, African American students have lower percentages proficient than their peers of other ethnic backgrounds. More specifically, from 2003-2006, the percent of African Americans proficient grew at a level equal to or less than the growth of higher achieving groups: 7 percent in English language arts and 5 percent in math, compared to growth rates for Asian (9% and 7%), Filipino (10% and 10%), and white (7% and 6%) students. In other words, the growth for African Americans is not at a rate fast enough to close the gaps.

Examination of data publicly available through the CDE Web site [27] shows that the percent of students proficient on CST tends to decrease by grade level across most ethnic groups and content areas, including African American students, with students at the elementary level showing higher proficiency rates than students in middle school and, in particular, high school. Thus, the gaps remain apparent across grades even though individual scores may change. For example, for test score year 2007, 66 percent of white students and 39 percent of African American students in grade 2 were proficient or above in English language arts, while only 55 percent and 23 percent achieved proficiency respectively at grade 10 [27].

Additional information can be gained from examining the most recent (2007) results available for students at the lowest levels of proficiency, that is, those in the “far below basic” category on the CST [27]. Across grades and content areas in math, language arts, and also science, African American and Hispanic students have the highest

Table 2

California 2006 base Academic Performance Index (API)

	All Grades	Grades 2-6	Grades 7-8	Grades 9-11
Overall	721	752	716	683
Subgroups				
Ethnic/Racial				
African American (not of Hispanic origin)	635	677	623	589
American Indian or Alaska Native	691	724	678	659
Asian	847	876	856	807
Filipino	808	846	806	763
Hispanic or Latino	656	690	644	612
Pacific Islander	714	761	704	661
White (not of Hispanic origin)	801	837	803	759
Socioeconomically Disadvantaged	654	686	640	607
English Learners	637	676	618	586
Students with Disabilities	518	567	499	456

Source: http://api.cde.ca.gov/AcntRpt2007/2006Base_StAPI.aspx

percentages of students in this group compared to other ethnic groups, with the percentage of "far below basic" African American students matching, and often exceeding, the percentage of Hispanic students.

Looking at the end of California students' K-12 educational experience, a similar pattern to those on the CST can be found across multiple years of the CAHSEE. Specifically, African American students have lower pass rates for both math and language arts across the years than white, Asian, or Filipino students [24,27].

Hispanic Students

The majority of the patterns in the achievement gap discussed for African American students are also applicable to Hispanic students. Looking at the API, the Hispanic subgroup score was lower than all ethnic groups other than African American for 2006 base and 2007 growth across grade levels (see Tables 2 and 3), with a slight decrease in the gap across the years. For example, in 2006 and 2007 there were 145 and 140 point gaps between Hispanic and white students, showing a small reduction trend.

Looking beyond the API at the CST scores across multiple years, Hispanic students have a lower percent proficient than all other ethnic groups for English Language Arts (ELA), and for all groups except African Americans in math (see Tables 4 and 5 and [12]). Growth over the years in the percent of Hispanic students proficient was equal to or less than other ethnic groups in ELA, and only higher than African American and white students (by one percentage point) in math, suggesting as with African American students that growth has not been at a faster rate than higher achieving groups (i.e., what would precipitate a closing of the gap).

Table 3

California 2007 growth Academic Performance Index (API)

	All Grades	Grades 2-6	Grades 7-8	Grades 9-11
Overall	728	761	720	689
Subgroups				
Ethnic/Racial				
African American (not of Hispanic origin)	643	688	629	596
American Indian or Alaska Native	696	727	688	664
Asian	852	880	861	814
Filipino	813	849	814	768
Hispanic or Latino	665	702	651	621
Pacific Islander	720	765	709	669
White (not of Hispanic origin)	805	840	807	765
Socioeconomically Disadvantaged	662	697	647	616
English Learners	646	690	623	590
Students with Disabilities	527	579	502	463

Source: http://api.cde.ca.gov/AcntRpt2007/2007GrthStAPI.aspx

Again, gaps between Hispanic students and higher achieving groups start at the earlier Grades, and persist through high school [28]. For example, in 2007 45 percent of grade 10 Hispanic students are below or far below basic on the CST in ELA, compared to 20 percent of white students (rates at grade 4 were 28 percent and 9 percent, respectively) [27].

CAHSEE pass rates for Hispanic students are similar to African American students in ELA, but slightly higher in math. For example, in 2006 approximately 50 percent of African American and Hispanic CAHSEE takers passed the ELA portion, while the numbers for math were 40 percent and 49 percent, respectively [24].

English Learners (ELs)

The assessment process for ELs is an area warranting its own in-depth analysis beyond the scope of this paper. For example, it has been suggested that the language demands of content area achievement tests may vary widely by content area and grade, with the specific academic language demands (both for general academic mastery and specific to the content area under considerations) of the tests of particular importance (see [29,30,31]). With that in mind, not surprisingly throughout the various California test outcomes ELs perform at a lower rate than non-ELs. There is also a persistent API difference based on EL status, with EL students performing below the state average both in 2006 and 2007 (see Tables 2 and 3). In viewing these tables it is important to note the overlap of EL status with other subgroups. For example, the vast majority of ELs in the state are Hispanic, and socioeconomic status also varies by ethnicity within the state so EL status also co-varies with economically disadvantaged students [23].

Table 4

Standardized Testing and Reporting (STAR) Program, California Standards Test results, percentage of students scoring at proficient and above by subgroup, 2003-2006, English language arts

Subgroup		2003	2004	2005	2006
All Students		35%	35%	40%	42%
Gender	Female	39%	40%	44%	46%
	Male	31%	32%	36%	38%
Ethnicity	African American	22%	23%	27%	29%
	American Indian/Alaskan Native	31%	31%	36%	37%
	Asian	55%	56%	62%	64%
	Filipino	48%	50%	55%	58%
	Hispanic/Latino	20%	21%	25%	27%
	Pacific Islander	31%	31%	36%	39%
	White	53%	54%	58%	60%
Economically Disadvantaged Students		20%	21%	25%	27%
Non-Economically Disadvantaged Students		49%	50%	56%	58%
Students Receiving Special Education Services*		9%	10%	11%	13%
Students with no Reported Disability		38%	38%	43%	45%
English Only Students		44%	44%	49%	51%
Initially-Fluent English Proficient (I-FEP)		46%	48%	53%	56%
Reclassified-Fluent English Proficient (R-FEP)		40%	42%	48%	50%
English Learners		10%	10%	12%	14%

* The percentages for Students Receiving Special Education Services do not include the results for students who were administered the California Alternate Performance Assessment (CAPA). Percentages included in this table may differ from the percentages printed on the internet reports due to rounding.

Source: http://www.cde.ca.gov/nr/ne/yr06/yr06rel89.asp

When looking specifically at individual CST scores (Tables 4 and 5) proficiency rates for ELs are not surprisingly lower for ELA than math, although there was some overall growth from 2003 to 2006 (10 percent to 14 percent and 20 percent to 25 percent in ELA and math, respectively). For ELs the differences between percent proficient at the lower and upper grades are particularly pronounced—in 2007, while 30 percent of grade 2 ELs are proficient or above in ELA and 46 percent in math, only 4 percent are proficient or above in ELA in grade 10, and 13 percent proficient or above in CST math at the grade 7 [27]. Needless to say, EL CAHSEE pass rates are lower than non-ELs, although rates are higher for math than ELA. For example, for the 2006 and 2007 administration, while passing rates for ELA ranged between 18 percent and 38 percent, depending on year and student grade level, passing rates for math were as high as 47 percent [24].

Table 5

Standardized Testing and Reporting (STAR) program, California Standards Test results, percentage of students scoring at proficient and above by subgroup, 2003-2006, mathematics

Subgroup		2003	2004	2005	2006
All Students		35%	34%	38%	40%
Gender	Female	34%	34%	38%	40%
	Male	35%	35%	39%	41%
Ethnicity	African American	19%	19%	23%	24%
	American Indian/Alaskan Native	29%	28%	32%	35%
	Asian	60%	60%	65%	67%
	Filipino	44%	45%	50%	54%
	Hispanic/Latino	23%	23%	27%	30%
	Pacific Islander	31%	31%	35%	38%
	White	47%	46%	51%	53%
Economically Disadvantaged Students		24%	25%	29%	30%
Non-Economically Disadvantaged Students		45%	44%	49%	52%
Students Receiving Special Education Services*		13%	13%	15%	16%
Students with no Reported Disability		37%	36%	41%	42%
English Only Students		39%	39%	43%	45%
Initially-Fluent English Proficient (I-FEP)		44%	45%	49%	52%
Reclassified-Fluent English Proficient (R-FEP)		37%	37%	41%	53%
English Learners		20%	20%	24%	25%

* The percentages for Students Receiving Special Education Services do not include the results for students who were administered the California Alternate Performance Assessment (CAPA). Percentages included in this table may differ from the percentages printed on the internet reports due to rounding.

Source: http://www.cde.ca.gov/nr/ne/yr06/yr06rel89.asp

Other Subgroups

Although certain patterns have emerged across the various California measures in terms of students who are at the low end of the achievement gap, it is also worth noting positive trends among other groups. Female students, for example, not only outperform male students in ELA (an area where women have traditionally scored higher than men), but now either match or exceed males' performance in math and science (areas where women have in the past performed lower than men [32]). Filipino students as a group also perform highly across all measures, even though this group initially faces its own unique set of language and cultural barriers in immigrating to the United States [33]. In terms of language status, redesignated students also perform highly across all measures. Although this warrants further study, it again suggests that EL students who stay in EL status for a longer term or who enter school at an advanced age are at particular risk when it comes to the achievement gap.

Kindergarten and Preschool

As CST testing begins at grade 2, the findings for younger students come from less uniform sources and are more limited. With that in mind, a recent review of the available evidence at the kindergarten and first-grade level, including California ELA curriculum end-of-year assessments and school readiness measures, found similar patterns in the early grades to what is shown in the CSTs [34]. For example, for both of the adopted California ELA programs' (end-of-year skills assessments) African American and Hispanic kindergarteners performed below white and Asian students during the 2004-05 and 2006-07 school years. ELs again performed below non-ELs. Some of the gaps at the elementary and secondary levels can be found at the preschool level as well, although again the findings are limited. For example, while only 47 percent of students age 3-4 in the state are enrolled in preschool, the enrollment percent varies by ethnic background, with Hispanic students (38%) less likely to be enrolled than white students (58%) [35].

Summary

Subgroup differences found nationally are evident at the state level as well. African American and Hispanic students achieve lower than other ethnic groups across multiple measures tracing from kindergarten through graduation, and although there are some examples of small decreases in the gaps over time they are not at a pace to see a closing of the gap in the foreseeable future. ELLs also continue to perform lower than their non-ELL counter parts, although there are variations within different groups of ELL students.

Findings from Higher Education

The roots of the achievement gap found in K-12 education predict evidence of disparity in post-secondary educational attainment as well. The complexities of achievement differences in higher education clearly warrant their own detailed analysis, but for the purposes of this paper we will review only some of the pertinent trends and issues. Looking broadly at national college and university enrollment, the percentage of college-age individuals currently enrolled in college or university varies widely by ethnic background [36]. Hispanics (24.7%) and African Americans (31.8%) trail whites (41.7%) in terms of the percentage of college age students enrolled, with Asians (60.3%) showing the highest percentage.

Examining enrollment distributions based on type of college provide additional information about differences in the post-secondary attainment of students of different ethnicities. Although nationally white students represent the largest percent of college students at all types of colleges and universities, the representation of minority students varies based on the type of college involved. For example, Asian students represent a greater percentage of the student body at Doctorate granting institutions than they do at other four-year institutions or community colleges. Hispanic students, on the other hand, comprise 14 percent of all students enrolled at community colleges, and only 6.4 percent of those at Doctoral granting institutions [37]. In other words, while overall Hispanic college enrollment trails that of other ethnic groups, they are particularly under-represented at the "top level" universities. The same patterns have been found when comparing 4-year college education graduation rates across ethnic groups as well (i.e., the percent of enrolled students who eventually obtain a degree), with Asian and white students graduating at

higher rates than Hispanic and African American students [38].

Specific to California, an examination of publicly available enrollment data from the California public university systems is an even more pronounced example of these national trends. While the percentage of white students enrolled in California state colleges is relatively consistent regarding the type of college, enrollments for the other ethnic groups vary considerably based on college type [39]. For example, based on 2006 enrollment statistics, Hispanic students represent 29 percent of community college enrollment, but only 23 percent of CSU enrollment and 13 percent of UC enrollment. Again, differences in attainment between groups grows more pronounced as the competitiveness of the schools increase, with Hispanic and African American students represented in small numbers at the highest tier universities. Women represent a higher percentage than men across all ethnic groups, consistent with national trends, although the experiences and pathways of these women likely differ based on ethnic background [40,41,42].

It is worth noting also that many students entering the California university system do not have minimum proficiency in math and language arts; for example, in 2006 only 63 percent of first year students at the CSU met basic proficiency in English, 55 percent in math, with numbers estimated up to 75 percent at some campuses [43,44].

The community college system, in California and nationally, provides transfer opportunities for students, both in terms of building on their content knowledge and broader learning skills, and also providing supported routes for transfer to state four-year colleges. Some studies suggest gaps in this area of attainment as well [45,46], with transfer rates varying based on ethnicity, and again Hispanic and African American students representing the lowest percentages. It may also be the case that some of the enrollment choices students make may help to perpetuate these differences in eventual four-year college transfer. For example, Hispanic and African American students also tended to trail in other predictors of transfer success, such as the percentage of four-year college transfer-eligible classes taken in the first semester and GPA in transfer-eligible classes.

Research on Gap Correlates and Reduction Strategies

The purpose of this section is to review some of the research regarding the factors associated with the achievement gap, both to provide a context for the findings described above and also as background for the recommendations we bring to bear in the final section of this paper. First, we will examine some of the school context characteristics that tend to co-vary with student groups on the lower end of the achievement gap. Second, we review the current research about factors associated with reduced achievement gaps at schools.

Achievement Gap Correlates

When examining the nature and extent of achievement gaps it is also important to be aware of the contextual factors that often co-vary with student background characteristics. Given the complex web of factors that interact in a student's educational experience, and the extent to which school demographics associated with achievement gaps (ethnicity, EL status, SES, urban vs. rural setting) co-vary [47,48,49,50] these contextual differences cannot be assumed to cause gaps in

achievement. However, they are important in providing the larger picture where the achievement gap sits. Research has shown that schools that enroll traditionally under-represented student groups differ systematically from those that do not, with some of the most common findings reviewed below.

Teacher Differences

One important set of contextual factors related to student background variables is teacher characteristics. Teachers at schools that serve traditionally underrepresented populations (minority students, low SES, ELLs) differ in terms of preparation and experiences from teachers at other schools. For example, high-poverty schools and schools that are higher in minority students have been found to have more teachers with less than five years' experience than low poverty schools do, and also the highest percentage of teachers teaching "out of field" (i.e., without a credential in the content area being taught) in language arts, math, and science [47,49,51]. Teachers at these schools are also more likely to change schools [36]. Both credential status and teachers' transience have been associated with lower academic achievement [52].

Resource Access/Availability

Researchers have found that schools differ systematically in other resources beyond teacher characteristics. For example, researchers have found that schools with higher percentages of underrepresented students have the highest overcrowding and are less likely to offer enough (a)-(g) (college prep) courses for all students [47]. Teachers at predominantly low SES schools are less likely to report that they always have access to the textbooks that they need, and both SES and minority status has been associated with higher rates of students coming to school without books and writing utensils [22,53]. Related to this, teachers reported tangible support available at their school as their top reason for teacher turnover. Students at schools serving underrepresented students also have less access to school counselors [47].

The above is by far not an exhaustive listing of school background characteristics associated with demographic or achievement differences, and also does not take into account the myriad of out-of-school factors alluded to at the outset of this paper, such as parent and home factors, that have also been associated with differences in academic outcomes (e.g., [54,55]). The discussion does serve to reinforce that the achievement gap is complex and multi-faceted, and associated with a tangled network of inter-related student, teacher, school, and community-level variables that make the identification of a cause or small list of contributing causes difficult [56].

Research on Achievement Gap Reduction

There is a large body of research, theory, and debate about the impact and effectiveness of different instructional practices in improving general student learning and achievement (e.g., [57]). A review of these practices and varying views regarding their effectiveness is beyond the scope of this paper. More to the point is research investigating factors that have been associated with reducing the achievement gaps [58,59,60,61,62,63,64,65,66]. These studies tend to be case study or correlational in nature. For example, they examine common practices and characteristics of school that have shown to be effective in reducing achievement gaps in regards to traditionally under-performing students or investigate factors that predict increased achievement among schools that predominantly serve these students. At present there is neither a broad enough nor

empirically rigorous enough research base about the most effective approaches to reduce achievement gaps to make generalizable conclusions. However, given these limitations, there is consistency across many of the existing studies about the commonalties among schools that have reduced the expected achievement gaps. These commonalties tend to fall into three general categories relating to the types of capital, or school-level resources, that they speak to: Social capital (inter-personal factors such as shared trust, beliefs, goals, values, and expectations); intellectual capital (shared knowledge, intelligence, and competencies) and organizational capital (issues relating to broader resources and their employment) [67].

Social Capital

Across the studies described above, the shared beliefs and expectations at a school emerged repeatedly as factors distinguishing schools that showed evidence of closing the gaps. These beliefs are marked by feeling of trust and collaboration among school stakeholders as well as an identifiable, shared vision/school mission that focuses on clear, measurable academic outcomes and goals. Academic expectations for the school and students are high within the school culture for all students regardless of their previous academic background. Teachers are willing to share and support each other towards the achievement of schools goals, as well as are open to feedback from their colleagues to improve their own practices. Families are part of these educational communities as well, with on-going connection and communication between teachers/administrators and parents about school vision, goals, progress, and outcomes.

Intellectual Capital

As with social capital there were several trends across the studies in terms of knowledge/competency resources at the exemplar schools. The school curriculum is coherent, consistent, and linked to the standards. Issues of teacher quality and knowledge are paramount throughout. Teachers have strong subject matter knowledge for the content they teach, as well as a clear understanding of the standards and their connection to instruction. The schools offer professional development, but it is focused, content based, and tied to the overall instructional curriculum. In other words, these schools just do not provide copious professional development activities, but rather target professional development purposely to support the academic outcomes and goals of the schools (rather than individual teacher professional interests). Another consistent trend is the regular use of assessment data to examine current practices and improve instructional (i.e., use assessment information for formative purposes rather than only summative purposes).

Other Organizational Resources

Although less prevalent than the social and intellectual factors described above, other types of resources emerged across the studies. Most obviously, the schools had adequate classroom instructional materials (i.e., the needed instructional and classroom resources were available to teachers). These schools also tend to develop community partners/resources and use these partnerships towards targeted ends (i.e., programs and activities designed to promote academic achievement). Wrap-around instructional support was accessible to students, including advisory programs, school counselors, after school and summer school programs, and early warning intervention activities (i.e., identifying students in need of extra help

early before problems may prove intractable).

Again, these number and design of these studies limit the conclusions to be drawn from these findings. The types of best practices factors identified are very often the same types of factors associated with quality practice at all schools, so the extent to which these factors will provide a quality learning environment for any student vs. working specifically to close gaps in achievement between different subgroups is not clear. Additional research with larger samples and more empirically rigorous methodology would be needed to further tease out the issues of what factors specifically function to close achievement gaps vs. those factors that approve achievement overall within a school for all student subgroups. In other words, as of yet there are no "silver bullets" with extensive empirical support for reducing a school's achievement gaps.

Measures

Part of our charge is to consider the standards and measurement system now in place and its role, if any, in reducing the gap. While strong in a number of ways, our judgment is that the present system is flawed if its principal intent is to provide information to improve the educational status of all students. We believe there are ways to repair or restructure it. In overview, we will address standards, existing tests, validity consequences, and detail problems that could be overcome.

Standards

We contend that one deficiency is that the system depends upon standards that are too numerous to teach well, that is, in depth. First, fewer standards, seriously be measured, can streamline and improve the system and guide learning [68]. These standards would need coherence that can be provided in part by including in content measures of cognitive readiness, such as, application, problem solving, and adaptation of learning. To have fewer standards, we need a careful selection process to avoid the idea of minimum competencies. Instead they should include important content, domain skill, and individual cognitive skills, such as self-monitoring, that might cross subject-matter boundaries. Fewer standards would allow the careful analysis of prerequisite or supporting skills, features that might be measured and shared with teachers for use in classroom formative assessment. Fewer standards would permit a more robust assessment system, one that not only measured annual skill acquisition, but retention, application, sustainment, and integration through periodic standard and non-standard procedures.

Testing Concerns

A system with as many standards as California's uses a survey test that by necessity must attempt to cover too much content with too few items per content area. For example, only 1 or 1/3 of an item addresses some sub-standards. With accountability sanctions such as those in NCLB legislation, and by various local board expectations, we all understand that pressure is placed on administrators and teachers to "get results" on the current set of tests. Suppose you were a teacher who took a lightly measured (i.e., few items covering it) sub-standard seriously and invested a significant proportion of instruction there, because, you felt based on research, theory, or experience that the learning was foundational for future learning. However, the teacher's effort would not be rewarded

on the test. Effort on those sub-standards would count little toward a good test score. Low item frequency signals a lack of importance of certain content and modifies the intention of standards. The system becomes test driven rather than standards driven, a satisfactory outcome if the test itself has depth and is connected to sustained and important learning, but not so in a survey test situation. The price of survey tests is relatively low, but their educational costs may be high, as opportunity costs for interesting, motivating instruction for students and teachers are levied. Furthermore, the schedules of events leave little time for even those areas where students need additional help to achieve prerequisite knowledge and skills. In other words: Fail once and continue to fail. Attention is given to moving numbers of students in the right classification, with less time available to promote deep and sustained learning for individual students. Teachers need to redirect their efforts more to the learning of individual children than to the servicing of an accountability system, and they can do this by using formative assessment. If teachers are given an empowered role in the assessment system, with clear guidance about intentions of standards, prerequisite skills, they can apply classroom formative assessments that give usable, significant diagnostic value. For example, CRESST has some preliminary evidence of a significant effect size with formative tests [69]. As an aside, much of what goes by the name of the buzz world "formative" does not function in this way, as it is neither diagnostic, provides no short turnover feedback, nor is it connected to timely and effective instructional options.

Validity, Now and the Future

While authorities may be satisfied by the "alignment" of tests and standards and content validity (e.g., there is a fragment, or one or more items that represent some aspect of the verbal standards), there is far less evidence that the current test results serve intended purposes. Many educational purposes are claimed for assessments, including monitoring effectiveness of individuals and schools, measuring progress, providing diagnostic information for instruction, supporting overall system improvement, and communicating with the public and with educators and teachers about important outcomes. Implicit in the test sequence is the notion that passing a test at one level should set a child up for doing well in the next, that is, there is a cumulative aspect to the measured content and cognitive demands. It is also assumed that the tests do not unintentionally include barriers that might result in unfair differential performance as a result of syntax, language, or content.

Yet, there is sparse validity evidence available about the California tests for most of these purposes. For example, and clearly relevant to achievement gaps, it must be demonstrated that performance on the CSTs can be affected by high-quality instruction (i.e., more than test practice) for students who have achieved necessary prerequisites, In other words, that what high quality schools do in classrooms shows up and overcomes as much as possible differences in background, setting, and other correlates of poor performance. If such empirically rigorous evidence is not available, then it is questionable to claim such measures determine school effectiveness, improve individual growth, and should be the basis to invoke sanctions. Whether the tests are sensitive to instruction [70] does not mean that one designs instruction solely to raise test scores. Rather, it means that the tests results will change because of instruction and that instruction will have a greater impact than irrelevant out-of-school activities, personal characteristics, and other

factors external to the system. Have there been clear studies that address this question in a way that disentangles student and school characteristics, teacher preparation, and a range of instructional options? Few or none.

Another purpose of tests is to mark the progress and status of schools and groups, and ultimately to allow the individual monitoring of student performance progress for the purpose of improvement. Is there evidence to date that data from tests can be used to develop high-quality instruction, or to improve or refine instruction? Are teachers and other educators using the standards and tests to diagnose and improve instruction? Has it been shown that teachers have a range of instructional tools at their disposal and can determine why students do poorly on certain standards or sub-standards? Is there evidence that fundamental sub-skills that precede the desired performance at particular grade levels are available for rapid instructional use? Are the test results timed and reported in such a way as to permit a serious effort? From the overall progress in California, we would guess not. Instead, there is some evidence that preparation for the test itself (i.e., for the formats and content of the subject) occupies the attention of schools at risk for failing to meet their API or AYP standards, an apparent national trend. Evidence of this sort of practice includes the selection of interim or benchmark tests intended to predict student performance, without budgeting instructional time to remedy needed skills or content gaps. Nor is it evident that tests in use have rigorous empirical support of their appropriateness for other possible purposes, such as supporting continuing progress in K-12 and beyond in the world of work or in further education We also mentioned the concept of cognitive readiness as measured by assessments (the acquisition of important content, skills, and propensities to adapt to future requirements and opportunities in education and work) earlier. California must bring its assessment system in line with changing expectations for cognitive readiness beyond K-12 education, especially by and for secondary students. Overall, any learning that is deliberately addressed in school must have a goal of transfer and generalization. Whatever the standard, students must be expected to demonstrate it in a different way, a different setting, varied constraints, and varied practicality than has been heretofore exhibited in our State tests. Unless students are taught for transfer, and unless generalizing and adapting to new situations is valued, students will continue to learn in a non-cumulative way. Can it be shown that students who do well on the test also can achieve well in similar tasks that require that they transfer their expertise to a new situation, formal testing or otherwise? Validity studies for this purpose and the other purposes outlined above for assessments measuring fewer standards would still need to be conducted, yet we believe that their results might be trusted in part because of the depth of their measurement and their inclusion of key prerequisites.

Frameworks

Implementation of standards and tests need glue to make them usable and coherent. It is no small irony that there are states that perform well on NAEP compared to California and have used the California frameworks as a basis for their system design. We believe that frameworks where curriculum is explicated, content is detailed and explained, and instructional options are provided along with formative assessment guidance are a key way to support the coherence of our proposed system. California should augment the smaller set of standards described above with strong state

frameworks to guide practice for teachers, parents, and students, and community support educators. These frameworks must be made available and clear, address standards, content, instruction, and classroom assessment. They will stimulate the availability of interesting instructional materials and technologies. With standards, they must form the basis for the design of instructional material, professional development, and a serious proportion of pre-service education, including subject matter preparation in colleges and universities. The design and implementation of frameworks provide a key way to operationalize shared expectations for learning that can be expected at every stage of the academic continuum.

Frameworks can also be used to motivate ongoing collaboration and connection between K-12 education and both two-year and four-year colleges. Agreement and understanding around shared frameworks could help to provide a better shared understanding between K-12 and higher education about the gaps in learning that many entry-level college students bring to both the two-year and four-year college experiences and how to address them at critical transition points in the process (high school to two-year college, two-year to four-year). The framework could also assist colleges in the development of some of the college- and department-level student learning outcomes required for accreditation purposes [71,72].

Toward a Revised, Blended Assessment System

Our testing system must address fairness for students who have underperformed, whether this involves cultural, linguistic, or motivational reasons. Yet, the balance between uniform reporting and individual needs, interests, and skills must be confronted and explored in ways that support individual learning. Clearly the uniform testing approach alone is yielding little to be proud of in California. Our discussion above leads us to consider instead a blended approach. For the early grades we must make sure that fundamentals are learned in an applied context, and that, as a child grows, there are further opportunities for choice. By the beginning of high school, students should be given multiple opportunities to demonstrate their expertise and transfer in different ways. End-of-course examinations in secondary school can provide a deep dose of coherence by linking standards, frameworks, syllabi, and assessments together with sufficient flexibility and time to deal with individual student needs. Later choices of certification options in secondary school might be expanded so that students could attain certifications in academic subjects (e.g., honors courses), AP classes, demonstrating specific applied skills (i.e., computer network manager), or by completing significant internships in areas of interest (i.e., environmental studies, graphic design). To accomplish this differentiation, schools would need to be aided by non-profits, colleges and universities, businesses, and other community-based entities. Instead of passing only an exit examination, students could have the additional requirement to show their expertise in two or three separate subject matters or thematic areas either through formal tests or personal evaluation.

Summary

To summarize, the evidence suggests that achievement gaps are pervasive and persistent at the national and state levels, with little suggestion that the gaps will close

in the foreseeable future. These gaps start early and are evident even at the highest levels of education. The causes of the achievement gaps are multiple and interactive, and there is limited empirical evidence about school-level changes that can be made to reduce them. However, what we know about best practices in measurement and assessment can help us devise approaches to addressing the gaps. Based on these considerations, we contend that education must change from a school or group enterprise to an individual learning endeavor. Schools should be addressing individual needs for acquisition of unlearned prior knowledge, child-by-child, not group-by-group, grade-by-grade, or in any other summary matter. School organization needs to monitor and support individual growth. To do so, schools must be given access, through technology, to options that allow students to learn on their own. This learning must have motivational, cognitive, and content excellence, in order to reward persistence with real competence. If we start and end with the students, their particular learning needs and aspirations, our processes must change. Innovations will start with changing the everyday experiences students have at school, by supporting teachers or inventing better instructional opportunities.

Student and teacher learning needs to become paramount in assessment, rather than subsumed to bureaucratic requirements. Starting with children involves the full range and type of student—those learning English, those high performing (not mutually exclusive), those new to the school or the State, those with particular needs. This approach will cause dislocation at the outset as it will represent a shift from current practice. But starting and ending with children should resonate with educators. "Coherence" will be a key underpinning of the system we are proposing. Coherence is different from more chiropractic notions of alignment, where objects like test items and bits of standards line up. Coherence operates more like a gravitational force that helps align elements without disturbing their own rotations or atmospheres. Coherence is something that grows from within a system rather than being imposed from the outside. Were individual learning valued and addressed coherently, the State would need to provide different levels of instructional choices and support for different students. Coherence would be expected among standards, syllabi and tests, teacher development, and consequent in-service development. In many countries, vendors build in this coherence because they provide assessments related to a framework, detailed syllabus, instructional options, included materials, and professional development.

Implicit in this coherent approach also is the importance of the formative uses of assessment, and clear links between multiple assessment outcomes, frameworks, standards, a wide range of instructional options. This approach to formative assessment assumes targeted support for teachers; information and guidance to help guide teachers in making instructional choices based on what they learn from the range of assessment results available to them [69,73].

For such a coherent network of multiple forms of assessment options and instructional resources to be implemented effectively would necessitate not only more experimental research about what types of instruction work best under what circumstances, but also establishing a systematic approach to identifying standards coverage across multiple instructional and assessment options. Such an approach, with the assistance and coordination of California's higher community education, could result in an integrated system for various grade levels and standards areas that also incorporated more options for schools

and teachers to choose from. For example, schools would be able to choose among vetted vendors, and choice for students, teachers, and parents would be an important component within and among standards areas.

The components of this discussion about measures, beginning with our thoughts on standards and assessments, leads us to consider a statewide system of assessment that: 1) requires pervasive rethinking of the standards, the instructional frameworks or syllabi that support them, and the tests that measure accomplishment, their quality, utility, validity, and fairness; 2) recognizes the importance of choice for children and youth, particularly as they approach adulthood; and 3) in many ways turns the system upside down, with an intense focus on children and youth rather than on the system intended to serve them.

Conclusions

The major conclusions of the report are stark. A first important conclusion from the data is that we are not succeeding now, nor are there harbingers of improvement on the horizon. California students fall well below national expectations and averages for achievement, which means they are in the lower reaches of world performance levels. In California, as nationally, membership in identifiable sub-groups remains a strong predictor of level of success on examinations; the gap endures. California is near the bottom ranks of the other states in this regard and, while we can argue for the uniqueness of our circumstances, they alone do not forgive our responsibility to educate all students well.

A second, inescapable conclusion is that, to the extent it occurs in the early grades, learning is neither cumulative, well sustained in middle and upper grades, nor supported, repaired or relevant enough to our youth to change their performance patterns. While potentially a test artifact, California displays the national trend that academic performance decreases as students rise in the grade structure of schools, a fact persisting in the face of the differential drop out of lower performing students.

Third, we conclude from analyses of California policies, that our educational policies have not been consistent or sufficiently coherent, and that any approaches to support coherence have been inadequately sustained, occasionally ill-chosen, and too remote from instruction and learning.

In summary, then, is the negative conclusion that California has not done enough to change the day-to-day learning experiences of students. We have not designed nor applied widely effective instructional systems in which teachers have flexibility and sufficient content and pedagogical knowledge to address each student's need. We have not given a coherent and sustained approach to instruction, and we have not been effective in our approaches for under-performing groups. If we have spent much time and energy on structural and governance issues over the last two decades, it is now essential to shift that focus to the classroom. We conclude an overhaul is needed of our thinking and acting on the matter of achievement gap and its measurement. We need to change our attention to the performance patterns of individual students—within groups, if necessary—but the individual must override thinking in great swaths about remedies for groups or institutions.

Counter Arguments to Our Conclusions

Counter arguments may be made based on different interpretations, sources of research and observation of practice, and analyses of feasibility. They include: 1) that the effects of changes in California are below the surface and will emerge soon; 2) that we have fine standards and tests and don't want to undo the processes that melded political and programmatic elements; 3) that based on immigration and the myriad languages and cultures represented in California by first, second, and third generation learners, English learning is sufficiently successful; 4) that the state policy can only address at top-down mandates for global recommendations; 5) that to truly understand coherence we need trusted ways to measure instructional practice that don't currently exist; 6) that we should not add or reorganize outcomes midstream, especially for students who are currently struggling; 7) that any individual level student monitoring system will help even if it relies on limited test formats; and 8) that NCLB requirements have to be dealt with.

Given our previous arguments, we naturally find these statements wanting, and return to our prior assertion that we cannot make minor changes in policy and practice if we want to address the dismal level of California educational effectiveness as measured by existing survey instruments.

Obstacles to Policy Implementation

Obstacles to revised policy implementation in the areas of measurement, accountability, assessment, and coherence derive from both political and technical factors. On the political side, the players, including the State Board, (with its naturally changing membership and leadership), numerous *ad hoc* task forces, the State Department of Education, the Office of the Secretary of Education, institutions of higher education, and professional organizations, have only fitfully worked together, and on more than one occasion have operated at distinct cross-purposes.

In other ways we are not efficient or sufficiently focused on programs and incentives to assure appropriate outcomes. In California, because of our recent history and growing diversity, educators and policymakers spend considerable time, for instance, at school district levels or for innovative major projects, negotiating with group representatives about decision and program components, sometimes with great good will, other times, more contentiously. As a result, the interval for innovation start-up is protracted compared with more homogeneous settings, and leaves less time, energy and mobilization for making deliberative programmatic choices intended to focus on student learning.

A view widely shared by the public, policy makers, and educators is that performance on the "test"—whether it is CAHSEE or CST—is understood as the key outcome of education. It swamps all other options, including unsought but potentially compelling evidence about students' ability to retain, sustain, apply, and transfer their learning as measured on the external examination. One consequence of this reality is the ever-growing focus of instruction on just the item types on the test, rolling over content options outside of the measured domains, and interesting subtopics within the standards topics. Our lower-performing students are, as a consequence, given little choice, and school for some means a focus on the test, rather than the various ways the standards could be applied and made useful, especially at the secondary level.

With regard to validity questions, including the urgent need to document the impact of

instruction on measures used to fundamentally evaluate the effectiveness of education, the State has not pushed vendors of examinations to provide strong evidence on information on the range of purposes intended for the examinations: accountability, accuracy, representativeness of standards, instructional sensitivity, diagnostic utility by teachers, and consequences overall for the educational standing and health of the schools. This is not at all unusual and a function of speed and cost of implementing assessments.

In addition, there remain great differences in interpreting—beyond the external tests—common expectations of performance for all groups. In secondary schools, these differences in expectations and instruction can be documented by looking at the proportion of non-credit bearing remedial coursework required of entering freshmen in some sectors of the higher educational system, where often freshmen with high enough performance for college acceptance (e.g., grade point average or grade in relevant courses) find themselves in remedial settings. One could anticipate how this assignment leads to college dropout and disaffection as well as greater time and cost to complete a course of study. Thus, the inhibitions between widespread and serious connections between secondary school and institutions of higher education must be overcome by rapid and intense engagement of the P-16 community if a serious understanding of the standards and their importance for achieving social justice and intellectual parity is to occur.

Recommendations

Based on our review of existing data, the research base, and the state of the art regarding assessment policy and practice, we formulated a series of assessment-focused policy recommendations directed at addressing the achievement gaps in California. For any of these recommendations to succeed will require not only expertise from a variety of research and applied fronts, collaboration/trust between a wide range of stakeholders, and resources, but also long-term administrative and policy support—political will—for the changes to take hold. Given the time-sensitive nature of the work ahead, for each recommendation we provide a suggested timeline for completion and approval.

Standards

Standards need to be reorganized, reduced, and include elements of cognitive readiness integrated with content to support their coherence and cumulative impact.

Action

Begin immediately by convening experts in cognition, subject areas, and pedagogy with guidance from states that have already initiated the process, using new technology to support the process. Engage stokeholds P-16 (including pre-school and higher education) in this process. Suggested completion and approval: September 2008.

Curriculum Frameworks

Curriculum frameworks should be resurrected or refined for all subject areas at a level of specificity to guide subtask instruction, provide options for teachers in addressing shortfalls, and support the coherence and fairness of the system.

Action

Begin immediately by convening subject matter, cognitive and pedagogical experts from across sectors. Review available frameworks and existing California

documents. Create and test formats and, by coordinating with Standards group, lag only somewhat behind the standards group the creation, trial, and fielding of developed documents. Completion and approval: June 2009.

The Measurement System

Design

A blended measurement system is immediately needed that includes uniform external measures of a limited set of standards with clear and sufficient content and cognitive sampling. This endeavor would also include a formative assessment system driven by teachers to assess the acquisition of sub-skills and prerequisite knowledge outlined in the curriculum frameworks. At the secondary level end-of-course examinations should be acquired, implemented, and made a part of accountability systems. These examinations should offer opportunities for student choice and motivation. To the extent possible, the State should plan for high school exit criteria that also include acquisition of desired "world of work" skills certified by businesses, non-profit organizations, or other providers. It should be expected that California will move rapidly to assessment situations that capitalize on technology as a means of determining deeper understanding and more complex learning. As these elements come on line, the emphasis on CAHSEE as a graduation requirement might diminish. In addition, organizational effectiveness measures related to the fair distribution of social and intellectual capital for all members of the learning community should be undertaken and included in assessments of educational progress and level. These measures, along with measures of instructional practices, will move the system away from test-focused learning to standards-based accomplishments focusing on both students and the larger school community.

Actions

Begin immediately developing end-of-course examinations for secondary school, examining options in the U.S. and abroad. As standards are clarified, design (with vendors assistance), try out and use preliminary quality criteria for assessments that are part of the system (e.g., of instructional sensitivity, accuracy, standards fidelity, and fairness). Following standards approval, revise CSTs accordingly with technical approaches to maintain trend. Completion and approval: January 2010.

The Measurement System

Validation

The purposes of measures, indices, and other external indicators of effectiveness must be clarified, and evidence obtained to document or to guide improvement on the extent to which the test results interpretations support the purposes [74]. Key among these purposes is instructional efficacy, and evidence of the value of instruction for preparing students for required examinations is a minimum criterion. Evidence with respect to the acquisition of the content intended by standards also needs to be collected through additional measures of retention, application and transfer. These measures may well be incorporated at key points in the emerging measurement system to provide guidance and incentives for more coherent, integrated instruction, including attention to social capital. All validity studies must consider the performance of identifiable subgroups and linguistic attributes of the examination. The studies should be conducted as soon as possible with existing and on drawing board measures. Actions: Identify critical validity studies for measures of learning and school

effectiveness budget and schedule for needed validity studies. Develop mechanisms and fund studies by June 2110.

Interaction Among Sectors

Collaboration among the sectors of the P-16 group, from governance, structural, political, standards, training, and instructional perspectives should be accelerated by general agreements and detailed connections to promote coherence, fairness, understanding, common expectations, and support for under-performing students. These agreements must include professional organizations and other powerful players on the educational scene. A clear set of common goals and short-term milestones should be adopted and followed. Actions: Select and convene membership, identify subcommittees relevant to standards, underperforming groups, frameworks, implementation, professional development and validation. Ongoing.

Students First

Inherent in the changes proposed is that the system must shift its attention to the classroom and in particular to the growth and development of each student, independent of background, focused on instructional needs. Teachers will in many situations need to be provided with new incentives that focus on individual rather than group progress. In the context of this individual focus, gaps should diminish and new skills should emerge if our aforementioned recommendations related to standards, frameworks, valid assessments, technology, student choice, and rights to social capital, are taken together, taken seriously, and implemented with care and speed. Action: Start immediately with a campaign to model and illustrate what a focus on individual learning would operationally mean. Create or adopt approaches from other successful venues for use. Focus on micro-interventions rather than complex, unwieldy or burdensome packages. Encourage exploration with results. Create incentive system that rewards attention to individual growth. Ongoing.

References

1. Fletcher, J. D. (2004, September). *Cognitive readiness: Preparing for the unexpected* (IDA document D-3061). Alexandria, VA: Institute for Defense Analysis.
2. Marks, G. N., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation, 12*(2), 105-128.
3. Orfield, G. & Lee, C. (2005). *Why segregation matters: Poverty and educational inequality*. Cambridge, MA: The Civil rights Project, Harvard University. Retrieved March 18, 2008, from: http://www.civilrightsproject.ucla.edu/research/deseg/Why_Segreg_Matters.pdf.
4. No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115 Stat. 1425 (2002).
5. First 5 LA. (2008). Available online at http://www.first5la.org/
6. Governor's Committee on Education Excellence (2007, November). *Students first: Renewing hope for California's future.* Sacramento, CA: Author.
7. Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., McMaken, J. (with Jucovy, L. Z.). (2007). *Making a difference in schools: The big brothers big sisters school-based mentoring impact study*. Philadelphia, PA: Public Private Ventures.
8. National Education Goals Panel. (1991). *Measuring progress toward the national education goals: Potential indicators and measurement strategies* (Compendium of interim resource group reports). Washington, DC: Author.
9. National Council on Education Standards and Testing. (1992). *Raising standards for American education: A report to Congress the Secretary of Education, the National Education Goals Panel, and the American people*. Washington, DC: U.S. Government Printing Office.
10. Improving America's Schools Act of 1994, Pub. L. No. 103-382, 108 Stat. 3518 (1994).
11. California Department of Education, Policy and Education Division. (2000). *Parent guide to the 1999 similar schools ranks based on the Academic Performance Index.* Sacramento, CA: Author.
12. O'Connell, J. (2006). *Summary of results. 2006 Standardized Testing and Reporting (STAR) Program*. Retrieved February 29, 2008, from http://www.cde.ca.gov/nr/ne/yr06/yr06rel89.asp.
13. Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32-38.
14. Bracey, G. W. (2005). Put out over PISA. *Phi Delta Kappan, 86*(10), 797-798.
15. Bracey, G. W. (2007). The first time 'everything changed.' *Phi Delta Kappan, 89*(2), 119-136.

16. Organisation for Economic Co-Operation and Development. (2005). *PISA 2003 data analysis manual: SAS® users*. Paris: Author. Available online at http://www.pisa.oecd.org/dataoecd/53/22/35014883.pdf

17. McGaw, B. (forthcoming). The role of the OECD in international comparative studies of achievement. *Assessment in Education: Principles, Policy & Practice*.

18. U.S. Department of Education, National Center for Education Statistics. (2007b). Trends in the achievement gaps in reading and mathematics. *The condition of education 2007* (NCES 2007-064, Indicator 14). Washington, DC: U.S. Government Printing Office. Retrieved February 29, 2008, from http://nces.ed.gov/programs/coe/2007/pdf/14_2007.pdf.

19. Phillips, G. W. (2007, November 14). *Chance favors the prepared mind: Mathematics and Science indicators for comparing states and nations*. Washington, DC: American Institute for Research.

20. Institute of Education Sciences. National Center for Education Statistics. (2008). *The NAEP database* (the Nation's Report Card). Washington, DC: Author. Available online at http://nces.ed.gov/nationsreportcard/tdw/database/.

21. The Nations Report Card. (2007). Available online at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007494; http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007496.

22. U.S. Department of Education, National Center for Education Statistics. (2007a). *The condition of education 2007* (NCES 2007-064). Washington, DC: U.S. Government Printing Office. Retrieved February 29, 2008, from http://nces.ed.gov/pubs2007/2007064.pdf

23. Reyes, B. I., Cheng, J., Currie, E., Frakes, D., Johnson, H. P., Macro, E. B., Reed, D., Signoret, J., & Spetz, J. (2001). *A portrait of race and ethnicity in California: An assessment of social and economic well-being*. San Francisco, CA: Public Policy Institute of California. Retrieved March 3, 2008, from http://www.ppic.org/content/pubs/report/R_201BRR.pdf.

24. Becker, D. E., & Watters, C. (2007, October 24). *Independent evaluation of the California High School Exit Examination (CAHSEE): 2007 evaluation report*. Alexandria, VA: Human Resources Research Organization (HumRRO).

25. Sable, J., & Gaviola, N. (2007). *Numbers and rates of public high school dropouts: School year 2004–05* (NCES 2008-305). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved February 29, 2008, from http://nces.ed.gov/pubs2008/2008305.pdf.

26. Barton, P. E. (2005). *One-third of a nation: Rising dropout rates and declining opportunities* (Policy Information Report). Princeton, NJ: Policy Evaluation and Research Center, Policy Information Center, Educational Testing Service.

27. California Department of Education. (2008). *DataQuest*. Available online from http://data1.cde.ca.gov/dataquest/.

28. The Education Trust - West (2006, August 15). *Achievement in California 2006: Small gains, growing gaps*. Oakland, CA: Author.

29. Abedi, J. (Ed.). (2007). *English language proficiency assessment in the nation: Current status and future practice.* Davis, CA: University of California, School of Education. Retrieved March 3, 2008, from http://education.ucdavis.edu/research/ELP_Report.pdf.

30. Wolf, M. K., Kao, J., Herman, J., Bachman, L. F., Bailey, A., Bachman, P. L., Farnsworth, T., & Chang, S. M. (2008a). *Issues in assessing English language learners: English language proficiency measures and accommodation uses: Literature review* (Part 1 of 3, CRESST Rep. 731). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

31. Wolf, M. K., Kao, J., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008b). *Issues in assessing English language learners: English language proficiency measures and accommodation uses: Practice review* (Part 2 of 3, CRESST Rep. 732). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

32. Gallagher, A. M., & Kaufman, J. C. (2005). *Gender Differences in mathematics: An integrative psychological approach*. New York: Cambridge University Press.

33. David, E. J. R., & Okazaki, S. (2006). The colonial mentality scale (CMS) for Filipino Americans: Scale construction and psychological implications. *Journal of Counseling Psychology, 53*(2), 241-252.

34. Cannon, J. S., & Karoly, L. A. (2007). *Who is ahead and who is behind? Gaps in school readiness and student achievement in the early grades for California's children.* Santa Monica, CA: RAND Corporation. Retrieved March 1, 2008, from http://www.rand.org/pubs/technical_reports/2007/RAND_TR537.pdf.

35. Children Now (2008). *California report card 2008*. *The state of the state's children*. Oakland, CA: Author. Retrieved February 29, 2008, from http://publications.childrennow.org/assets/pdf/policy/rc08/ca-rc-2008.pdf.

36. U.S. Department of Education, National Center for Education Statistics. (2005). Minority student enrollments. *The condition of education 2005* (NCES 2005-094, Indicator 31). Washington, DC: U.S. Government Printing Office. Retrieved February 29, 2008, from http://nces.ed.gov/programs/coe/2005/pdf/31_2005.pdf.

37. U.S. Department of Education, National Center for Education Statistics. (2005). Student reading and mathematics performance in public schools by urbanicity. *The condition of education 2005* (NCES 2005-094, Indicator 14). Washington, DC: U.S. Government Printing Office. Retrieved February 29, 2008, from http://nces.ed.gov/programs/coe/2005/pdf/14_2005.pdf.

38. KewalRamani, A., Gilbertson, L., Fox, M. A., & Provasnik, S. (2007). Status and trends in the education of racial and ethnic minorities (NCES 2007-039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

39. California Postsecondary Education Commission. (2008). *Custom data reports*.

Available online from http://www.cpec.ca.gov/OnLineData/OnLineData.asp.

40. Constantine, M. G., & Greer, T. M. (2003). Personal, academic, and career counseling of African American women in college settings. *New Directions for Student Services, 104*, 41-51.

41. Sy, S. R. (2006). Family and work influences on the transition to college among Latina adolescents. *Hispanic Journal of Behavioral Sciences, 28*(3), 368-386.

42. Zarate, M. E., & Gallimore, R. (2005). Gender differences in factors leading to college enrollment: A longitudinal analysis of Latina and Latino students. *Harvard Educational Review, 75*(4), 383-408.

43. California State University (CSU). (2007). *CSU fall 2006 regularly admitted first-time freshman remediation: Entry level mathematics test*. Sacramento, CA: Author.

44. Shearer, J. (2007, June 25). CSUB expecting biggest freshman class. CSU The California State University, Daily News Clips. *Bakersfield California*, available online at http://www.calstate.edu/pa/clips2007/june/25june/big.shtml.

45. Driscoll, A. K. (2007). *Beyond access: How the first semester matters for community college students' aspirations and persistence*. Berkeley, CA: Policy Analysis for California Education.

46. Sengupta, R., & Jepsen, C. (2006). California's community college students. *California Counts Population Trends and Profiles, 8*(2), 1-24. Retrieved February 29, 2008, from http://www.ppic.org/content/pubs/cacounts/CC_1106RSCC.pdf.

47. Rogers, J., Oakes, J., Fanelli, S., Medina, D., Vallandares, S., & Terriquez, V. (2007). *California educational opportunity report 2007*. Los Angeles: University of California, Institute for Democracy, Education, & Access, and All Campus Consortium on Research for Diversity.

48. U.S. Department of Education, National Center for Education Statistics. (2003). *The condition of education 2003* (NCES 2003-067). Washington, DC: U.S. Government Printing Office.

49. U.S. Department of Education, National Center for Education Statistics. (2006). Postsecondary expectations of 12th-graders. *The condition of education 2006* (NCES 2006-071, Indicator 23). Washington, DC: U.S. Government Printing Office. Retrieved February 29, 2008, from http://nces.ed.gov/programs/coe/2006/pdf/23_2006.pdf.

50. Van Hook, J. (2002). Immigration and African American educational opportunity: The transformation of minority schools. *Sociology of Education, 75*(2), 169-189.

51. U.S. Department of Education, National Center for Education Statistics. (2005). *The condition of education 2005* (NCES 2005-094). Washington, DC: U.S. Government Printing Office. Retrieved February 29, 2008, from http://nces.ed.gov/pubs2005/2005094.pdf.

52. Fetler, M. (2001). Student mathematics achievement test scores, dropout rates, and teacher characteristics. *Teacher Education Quarterly, 28*(1), 151-168.

53. Oakes, J., & Saunders, M. (2002). *Access to textbooks, instructional materials,*

equipment, and technology: Inadequacy and inequality in California's public schools (Williams Watch Series: Investigating the Claims of Williams v. State of California). Los Angeles: University of California, Institute for Democracy, Education, & Access.

54. Barton, P. E. (2003, October). *Parsing the achievement gap: Baselines for tracking progress classroom* (Policy Information Report). Princeton, NJ: Educational Testing Service.

55. Kober, N. (2001, April). *It takes more than testing: Closing the achievement gap*. A report of the Center on Education Policy. Washington, DC: Center on Education Policy.

56. The Education Alliance. (2004, Spring). *Closing the achievement gap: Policy implications for teach quality, curriculum and teacher expectations*. Charleston, WV: The Education Alliance.

57. National Research Council (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: Commission on Behavioral and Social Sciences and Education.

58. Austin, G., Hanson, T., Bono, G., & Cheng, Z. (2007). *The Achievement gap, school well-being, and learning supports.* CHKS Factsheet #8. Los Alamitos: WestEd.

59. Bennett, A., Bridglall, B. L., Cauce, A., Everson, H. T., Gordon, E. W., & Lee, C. D. et al. (2004). *All students reaching the top: Strategies for closing academic achievement gaps*. Naperville, IL: Learning Point Associates.

60. Cohen, D. K., & Loewenberg Ball, D. (2001, September). Making change: Instruction and its improvement. *Phi Delta Kappan, 83*(1), 73-77.

61. EdSource Report (2007, September). *Similar English learner students, different results: Why do some schools do better? A Summary of a Research Study*. Mountain View, CA: Author.

62. The Education Trust (2005, November). *Gaining traction, gaining ground: How some high schools accelerate learning for struggling students*. Washington, DC: Author.

63. Henderson, A. T., & Mapp, K. L. (2002). *A new wave of evidence: The impact of school, family, and community connections on student achievement*. Austin, TX: National Center for Family & Community Connections with Schools.

64. Kannapel, P. J. & Clements, S. K. (with Taylor, D. & Hibpshman, T.). (2005, February). *Inside the black box of high-performing high-poverty schools: A report from the Prichard Committee for Academic Excellence*. Lexington, KY: Prichard Committee for Academic Excellence.

65. U.S. Department of Education, Office of Innovation and Improvement (2006). *Charter high schools: Closing the achievement gap, innovations in education*. Washington, DC: Author.

66. Williams, T., Kirst, M., Woody, E., Levin, J., Perry, M., & Haertel, E. et al. (2005). *Similar students, different results: Why do some schools do better? A large-scale survey of California elementary schools serving low-income students*. Mountain View, CA: EdSource.

67. Huang, D., Miyoshi, J., La Torre, D., Marshall, A., Perez, P., & Peterson, C. (2007). *Exploring the intellectual, social and organizational capitals at LA's BEST* (CSE Rep. 714). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

68. Popham, W. J. (Chair). (2001). *Building tests to support instruction and accountability: A guide for policymakers*. The Commission on Instructionally Supportive Assessment. American Association of School Administrators. National Association of Elementary School Principals. National Association of Secondary School Principals. National Education Association. National Middle School Association. Washington, DC: National Education Association. Retrieved March 4, 2008, from http://www.nea.org/accountability/buildingtests.html.

69. Phelan, J., Niemi, D., Choi, K., & Vendlinski, T. (2008, March). Some aspects of the technical quality of formative assessments in middle school mathematics. Paper presented at the annual meeting of the American Educational Research Association, New York.

70. Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, *20*(8), 15-21. (ERIC Document Reproduction Service No. EJ 436 999)

71. Wergin, J. (2005). Higher Education: Waking up to the Importance of Accreditation. *Change, 37*, 30-33.

72. Western Association for Schools and Colleges (WASC) (2001, January). *Handbook of accreditation*. Alameda, CA: Author.

73. Baker, E. L. (2008, March). *Empirically determining the instructional sensitivity of an accountability test*. Paper presented at the annual meeting of the American Educational Research Association, New York.

74. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.