

## Rejoinder: Being Fair to NCLB

David Rogosa  
Stanford University  
February 9, 2004

The functioning of NCLB accountability can be described as "no subgroup left behind" because in NCLB a school is evaluated on the basis of its lowest-scoring subgroup not on the average achievement in the school (min not mean). Thus, in NCLB the probability of failing to meet AMO is driven by the lowest achieving subgroup (regardless of whether there are few or many subgroups). By design, two schools with the same mean achievement can have different probabilities of satisfying NCLB. It would seem a hard sell to bash a program that does exactly what it sets out to do, but that's what is indicated in the title of the PACE Policy Brief, *Penalizing Diverse Schools? Similar test scores, but different students, bring federal sanctions*. In California, the PSAA accountability did function in the same manner as NCLB, but was based instead on year-to-year improvement in school and subgroup API scores (and with less onerous subgroup definitions).

The PACE Policy Brief (and most clearly in the accompanying December 23 UC Berkeley Press Release and December 23 *San Francisco Chronicle* story) recommended eliminating serious use of subgroup criteria in NCLB (e.g., reporting only). Without subgroup criteria for academic performance, a school of 200 students, 100 white and 100 African-American, could easily satisfy California AYP by having half the white students score proficient while none of the African-American students score proficient. No one would regard that pattern as acceptable. All policies have their pros and cons, and the subgroup criteria are no exception. I did the series of calculations in my Rebuttal to provide some quantification of the effects of the subgroup criteria on meeting AMO.

As subgroup criteria are a key part of the present NCLB and past PSAA accountability systems, it's important to understand the resulting statistical properties, and this discussion of the PACE Policy Brief provides an important opportunity. My Rebuttal was a teaching exercise with the PACE Policy Brief as a starting point, an attempt to explain basic statistical calculations for the NCLB and API contexts, bringing in other work similar to the PACE Policy Brief. (An attempt was made to constrain mathematical tools within the scope of the K-12 California Mathematics Content Standards with computational details and tools in the Appendix.)

One starting point for the probability calculations is provided by the December 23 UC Berkeley Press Release on the PACE Policy Brief in which Bruce Fuller explains the negative effects of multiple subgroups by saying: "It's simply more difficult to flip a silver dollar and get three tails in a row than to get one". Fuller's analogy is powerful, but one that is misplaced in NCLB discussion. Among the results in my Rebuttal was that the marginal effects of multiple subgroups are often rather small; if a diverse school (multiple subgroups) had the same level of underlying educational attainment (true proportion proficient) that would be required for a school having no distinct subgroups to have high probability of meeting AMO, then the diverse school would also have high probability for school and subgroups meeting AMO.

Importantly, the Novak-Fuller Response employs different language in describing the California NCLB results than was seen in the December Policy Brief. For example, no version of "unfairness" appears in the Response, and they seem to position the Policy Brief as simply describing disparities ("basic statistics") in AYP across schools, indicating in point 2 that they "simply raise the question". If the PACE Policy Brief had used the language and acknowledged the limitations as seen in this Response (points 1 and 2), then I would have not prepared a rebuttal. The PACE Policy Brief, December 2003, speaks extensively to "unfairness" of the NCLB subgroup criteria, attempts to empirically demonstrate a large "diversity penalty" and recommends eliminating subgroup criteria for meeting AYP (as also in the Dec 23 press release: "policy makers [should] improve accountability rules by eliminating the diversity penalty by rewarding schools when overall student achievement is growing, while still reporting on each subgroup's progress"). The assertions of unfairness and demonstration of a large diversity penalty generated the news coverage and thus the need for refutation.

The PACE Response does provide a welcome opportunity to explain again some of those probability calculations. In their point 5, the PACE authors use results from my Table 3 (upper frame, English) to determine a diversity penalty between schools with no distinct subgroups and four non-overlapping subgroups to be .057 in terms of required true proportion proficient. And this quantity is claimed to be similar to the PACE empirical findings. For one thing remember that 4 subgroups in the PACE analysis is equivalent to most likely 2, maybe 3, non-overlapping subgroups; a school with the equivalent of four non-overlapping subgroups is unusual. But more consequentially, their point 5 misinterprets an extreme result from what I describe there to be the worst case scenario (where number of subgroups has the greatest effect). In Table 3 school size is held constant as the number of subgroups increases, although in reality schools with

larger numbers of subgroups do not necessarily have proportionately smaller subgroups. Table 3 is presented as a bookend to Table 1 (in which subgroup size is constant as number of subgroups increases) in bracketing reality.

The comparison chosen in the PACE point 5, is between a school with 600 students, with the only subgroup composed of those same 600 students (e.g. all White school with no disadvantaged subgroup) and a school of 600 students with 4 non-overlapping subgroups each of size 150 students. The values cited by the PACE Response are .171 for the single group and .228 for 4 subgroups (in terms of school-wide true proportion proficient required for probability .99 that the school and all subgroups meet the English AMO). That increase of .057 in the required true proportion proficient would appear a substantial penalty for the multiple subgroups. But most of that increase is the effect of smaller sized subgroups (150 vs 600) not the effect of multiple subgroups. Using my Table 3 more carefully, a key entry is .211 for school of size 150 composed entirely of a single subgroup (probability .99 of meeting AMO with true proportion proficient .211). Thus over 70% of the effect cited in PACE point 5 is due to small subgroup size (.171 to .211), not to the "diversity penalty" of multiple subgroups.

The alternative comparisons keeping subgroup size constant are more stark. A single subgroup of size 600 versus a school with 4 non-overlapping subgroups of size 600 indicates the effect of multiple subgroups is less than one-seventh as large as the effect of subgroup size (as the increase in true proportion proficient between the single subgroup and four subgroup school is merely .171 to .178). Or from Table 1, comparing a school composed of a single subgroup of size 150 versus a school with 4 non-overlapping subgroups of size 150 indicates the effect of multiple subgroups is to lower the probability of meeting the English AMO from .99 to .961.

It is useful for Novak-Fuller to stress that my calculations don't have anything to say about the consequences of the participation rate criteria, employed in AYP in addition to the academic performance (AMO) criteria. I set that aspect of NCLB aside, as the participation criteria just aren't amenable to inclusion in the calculations for statistical uncertainty in the school and subgroup proportion proficient. Participation rate is a good topic for policy criticism, and PACE should have analyzed that aspect of NCLB in more detail. I will happily assert my view that the NCLB invocation of participation rate criteria in AYP determination is nonsensical and counterproductive. Missing students should be designated as having not demonstrated proficiency, and thus the missing scores will effect reported proportion proficient. If, as PACE authors indicate in their point 3, the participation criteria are the consequential factor in more diverse

schools failing AYP, then the proper action is to speak against the separate participation rate criteria, not to recommend against the serious use of subgroup academic achievement.

That all said, the fact remains that PACE Policy Brief does overstate the effect of subgroups. In particular, I call upon PACE to retract the 28% claim in their Key Findings and, most importantly, their policy recommendation that NCLB abandon serious criteria for subgroup academic performance. I am in complete accord with the PACE recommendations that states be given more authority in the formulation of the school accountability system; in many ways NCLB accountability is clearly inferior to the former California system based on improvement in school and subgroup API.

---

#### Additional Comments on PACE Point 4.

This fiction of an "advance copy" of the PACE Policy Brief is deeply troubling. Policy research dissemination via "do-overs" is not tenable. On December 23 2003 PACE released the Novak-Fuller Policy Brief, along with a UC Berkeley Press Release (by Kathleen Maclay), accompanied by a press blitz which produced Dec 23 stories in the New York Times ("New Education Law Is Faulted in Gauging Performance" By Sam Dillon) and San Francisco Chronicle ("Standards tougher on diverse schools" by Nanette Asimov). A week later I became aware of the PACE report and downloaded the copy I worked from. My January 7 Rebuttal did demonstrate some embarrassing blunders in the data display and exposition of the PACE Policy Brief, and sometime after I posted my document (PACE fully informed) the PACE authors saw the need for corrections to their Dec. 23 document. Clearly identified Errata are the proper mechanism for correcting a released document. Surreptitious amendment of a policy research document with national visibility would appear to damage the credibility of all PACE publications.

I am a credentialed mathematics teacher (though not in California), and therefore I did feel it necessary to protest PACE's interchangeable use of odds and proportions, as  $p$  does not equal  $p/(1 - p)$  except at 0. In this Response, the PACE authors reaffirm this interchangeability, and thus I will abandon my attempts to defend mathematics education. I feel obliged, however to point interested readers to the Mathematics Content Standards for California Public Schools page ([www.cde.ca.gov/standards/math](http://www.cde.ca.gov/standards/math)) with direction to the grade six and grade seven standards.