



219 North Main Street, Suite 402, Barre, VT 05641
(p) 802-479-1030 | (f) 802-479-1835

MEMORANDUM

TO: SBAC Governing States
FROM: Rebecca Holcombe, Vermont Secretary of Education
SUBJECT: Limitations of performance categories for supporting improvement in learning and assessing school effectiveness
DATE: November 2, 2014

Our purpose in meeting today is to approve performance thresholds on the SBAC assessments in compliance with the federal requirement that States report on student performance in terms of performance categories and review the “percent proficient.” Vermont does not support the use of performance thresholds as a valid means of communicating about performance. In addition, as of yet we have little empirical evidence related to the validity of the proposed cut scores for actually discriminating between those who are “college and career ready” and those who are not.

Our concern is all the more relevant given the high stakes and sanctions attached to performance relative to these threshold scores under federal policy. This is all the more worrisome given extensive research on the effect of persistent inequality on school outcomes. Since Vermont schools, despite being high scoring relative to the nation, are already considered “low performing” under NCLB, this will have little material impact in Vermont. However, I am sure it is a concern for other states at the table, where schools will predictably fail due to problems that are bigger than the tremendous efforts their teachers are making to improve the learning of our children. As Gary Orfield recently wrote:

“Setting absurd standards and then announcing massive failures has undermined public support for public schools. . . . We are dismantling public school systems whose problems are basically the problems of racial and economic polarization, segregation and economic disinvestment.” (Educational Researcher, August/September 2014, p.286)

Instead of reporting in terms of performance categories, we could report performance in terms of scale scores. A solid body of empirical research suggests that scale scores provide more complete information on performance and are more useful for the purpose of informing improvement efforts. We are indebted in particular to the

research of Andrew Ho¹ of the Harvard Graduate School of Education in coming to these conclusions:

1. **Cut scores are arbitrary:** Cut scores are set through human judgment in a standard setting process, and are therefore subjective determinations of “appropriate” levels of performance. SBAC is following best practice to set these scores, but there is no truly objective way to set threshold scores. As SBAC tests are new, these performance levels are not yet predictive of future outcomes. In fact, by design, SBAC is a test of mastery of established curricular standards. We can only hypothesize that it also captures elements of cognitive and other skills that will be useful in communities, college and in careers.
2. **Proficiency bands reward schools who push students over band levels and provide no recognition for growth within bands:** Any analyses of trend or magnitude of score gaps depend on where the proficiency thresholds sit relative to the distribution of test scores. For example, a school whose students initially score far below proficient might have very large gains in mean scale scores, but if the students are still scoring far below the threshold for proficiency, these remarkable gains will not be captured by reporting in performance categories. In contrast, a school that demonstrates very modest gains (perhaps one point average scale score) may see a very large increase in the number of students scoring as proficient, if its students, on average, are scoring very close to the proficiency threshold to start. In this case, the second school would appear to be more effective, while in truth the first is the more effective school. Thus, reporting in performance categories distorts and misrepresents the true story of improvement.
3. **Comparisons of the performance of subgroups (e.g. students in poverty vs. more affluent students) are distorted** in the same way comparisons in the performance of schools are distorted.

Reporting in performance categories has a public appeal: We like the apparent certainty and clarity of being able to categorize our students as “proficient.” However, this misrepresents the underlying complexity of achievement and contributes to simplistic policies that make it difficult to achieve our public purposes.

Vermont is present at this exercise because we are compelled to comply with threshold reporting by NCLB. However, the focus of state-level reporting in Vermont will be scale scores, and in particular, mean scores and changes in individual and mean scores. We believe this will support more responsible use of test data to inform improvement efforts at the school, district and supervisory union level. We encourage our SBAC

¹ Ho, Andrew Dean. (2008). The problem with proficiency: Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*. 37, 6, p. 351.

partners to also report using scale scores at the state level, so that we can collectively emphasize the commitment we share with respect to constructive use of assessments to support systematic improvements in learning.

We also invite our SBAC partners to join us in transparent discussions about what we currently can and cannot infer about students' "college and career readiness" based on SBAC scores.

#####