



Research Brief

Assessment, Research, and Data Analysis

Vol. 1302
October 2013

Dr. Terry Froman

Performance by Subgroup: the Sum of the Parts is Greater than the Whole

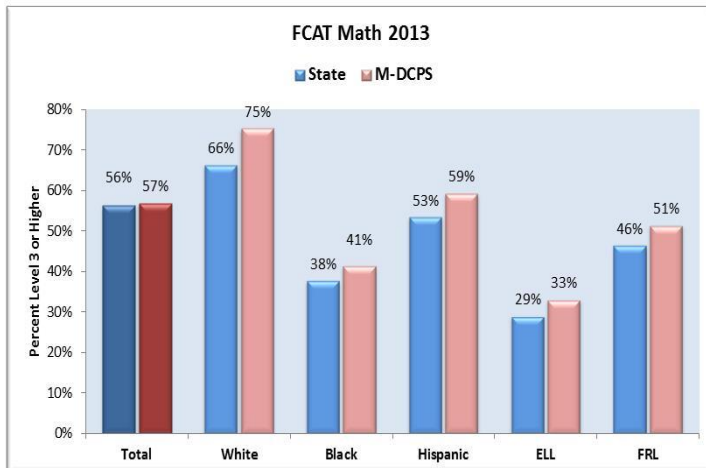
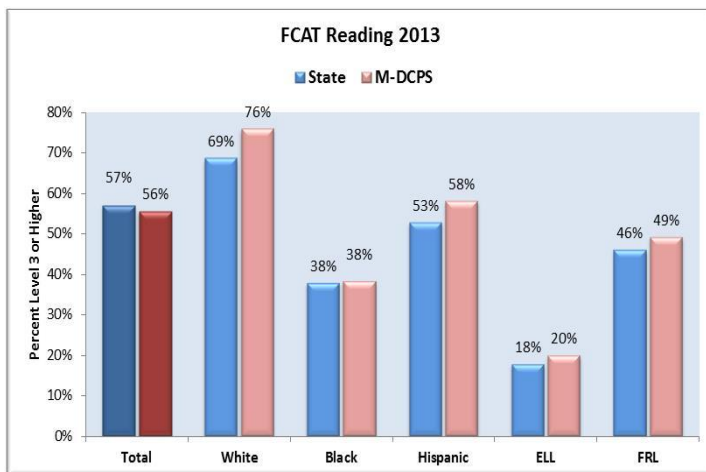
When comparing **overall** FCAT performance of the District to the State, the District comes up a little short in Reading and barely higher in Math. However, when the comparisons to the State are confined **within the typical subgroup categories**, the District meets or exceeds the State's performance in all subgroups.

This reversal in trend between the whole and the parts seems to defy common sense. How is it possible that we appear better than the State in all subcategories but worse overall? This is particularly perplexing when the combined subcategories comprise virtually the entire

population (as with White, Black, and Hispanic for our District). Which type of comparisons should be considered most fair? Where should we look to find reliable assessments of performance for the District?

The graphs to the left present the real data for students scoring at Level 3 or higher in Reading and Math FCAT for the 2012-13 school year across all grades. The Total category (far left) depicts the typical overall report for our District. All of the other subgroup comparisons in the graphs reveal a more accurate state of affairs: The District outperforms the State in all subcategories.

This research brief attempts to explain how the incongruity between overall and part comparisons can take place and underlines the importance of considering subgroup breakdowns when assessing district performance.

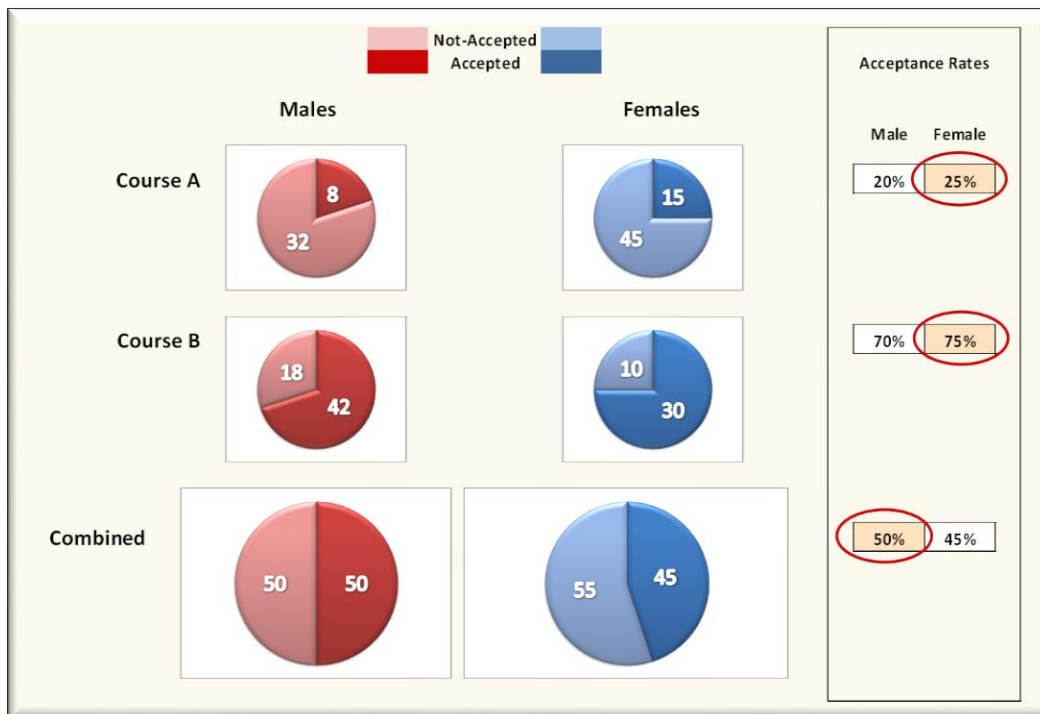


An Example of the Reversal Paradox

In 1973, the University of California-Berkeley was sued for sex discrimination. The numbers looked pretty incriminating: the graduate schools overall had accepted a greater proportion of male applicants. When researchers looked at the evidence, though, they uncovered something surprising: "If the data are **properly pooled**...there is a small but statistically significant bias in favor of women." (*Sex Bias in Graduate Admissions: Data from Berkeley*, P. J. Bickel, E. A. Hammel, J. W. O'Connell, *Science, New Series*, Vol. 187, No. 4175 Feb. 7, 1975, pp. 403)

When combining data from subgroups, it sometimes happens that apparent trends within subgroups reverse direction for the aggregate data. Although relatively unknown among laypeople, this kind of reversal paradox has been well-known among statisticians for some time. However, even among the well informed, this occurrence can be surprising and seem counter-intuitive. A simple example can help demonstrate how this phenomenon can occur.

Suppose there are only two courses for which male and female students are applying. The numbers being accepted and not accepted for each gender and their respective acceptance rates are displayed below.

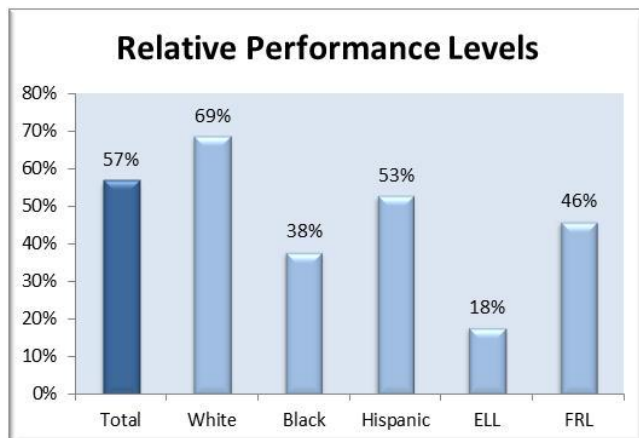


We can see in these graphs that females are accepted at a higher rate in each of Course A and Course B, but when we combine the data for both courses it appears males are accepted at a higher rate. This is true despite equal sample sizes of males and females overall and equal numbers of applicants for each course.

A closer look reveals the true source of the reversal paradox. It is evidently more difficult to gain acceptance into Course A than Course B. Additionally, a greater proportion of the male applicants as opposed to the proportion of female applicants (60% to 40%) attempt entrance into the course with the higher chance of acceptance. With a greater proportion applying to the easier-acceptance course, the males end up having a slight overall advantage.

So, if we are investigating whether there is gender bias with respect to course acceptance, should we pay attention to the overall rate (favoring males) or the rates within each course (favoring females)? In this example, and many others of the type discussed in this paper, the “fairer” comparison is made within the component subgroup. In each of the courses, with their widely different acceptance probabilities, the females are accepted at a slightly higher rate. This subclass comparison provides a more accurate picture of any potential bias. The overall rate is distorted in this example by the circumstance of *unequal proportions of subgroups in asymmetrical reporting categories*.

FCAT Performance Differences by Subgroup



In the graph to the left we have the percentages of students scoring at Level 3 or above for the 2013 Reading FCAT across the entire State, broken down by the common major subgroups. Here we can see that, relative to the Total population on the far left, White, Non-Hispanics performed at a higher level than did Hispanics, who, in turn, performed at a higher level than did Blacks. We can also see the characteristic lower levels of performance of English Language Learners (ELL) and students receiving Free or Reduced Lunch (FRL).

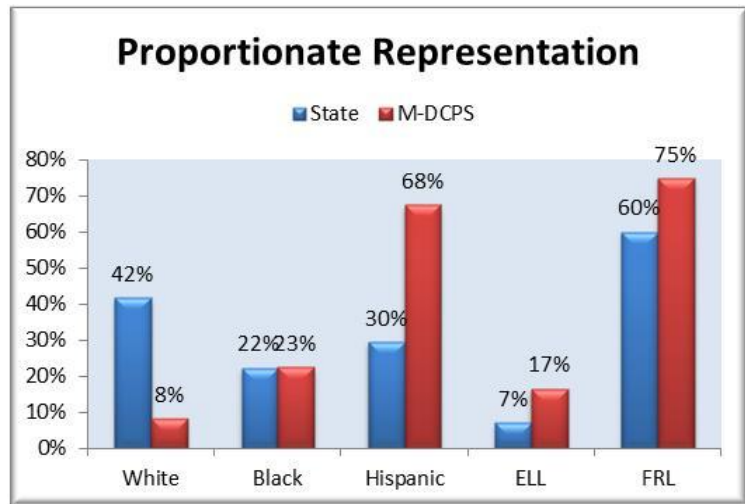
It is an observable fact that these subgroup performance differences remain relatively constant from district to district. The specific causes of these subgroup differences are hard to pinpoint, but presumably involve language difficulties, poverty issues, and subcultural influences. While differences in subgroup performance may be of genuine concern, the directions of those differences are universal in the state and not unique to our District.

Whatever the explanation, if the subgroup performance differences are parallel within each district, overall district summary data could be highly impacted by the particular proportions of each subgroup comprising a district. That is, if a district had a larger percentage of a lower than average performing subgroup, its overall performance average would appear lower when compared to the whole state. Likewise, if a district had a larger percentage of a higher than average performing subgroup, its overall performance average would appear higher. This would be true even if each subgroup performed exactly at its expected level consistently across all districts. Whether the overall average properly reflects the district performance is dependent upon the relative proportions of the different performing subgroups that comprise the district.

Disproportionate Subgroup Representation

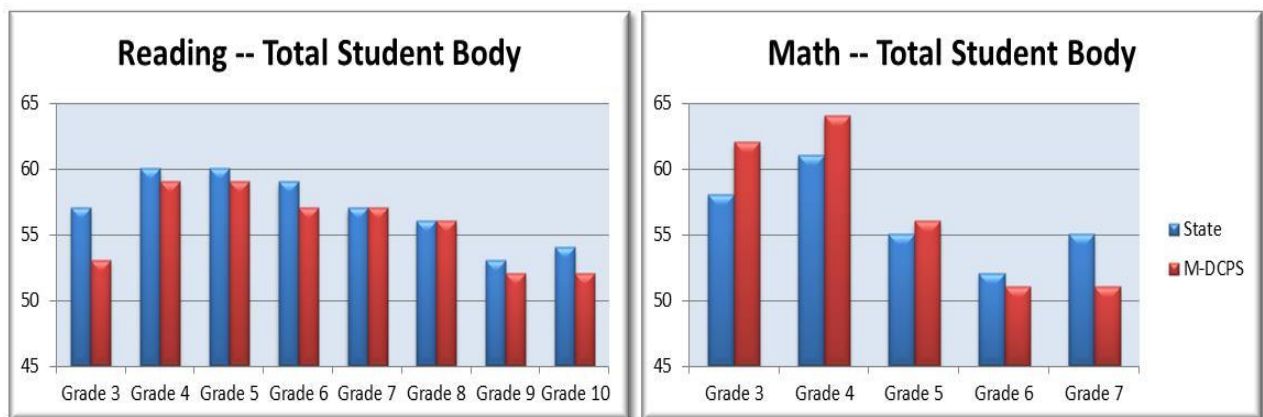
In this graph we can see the large discrepancies among the subgroup representations between the District and the State. There is a considerably smaller percentage of White students and greater percentages of Black, Hispanic, ELL and FRL students in the Miami-Dade District versus the State as a whole.

It should be obvious that lower proportions of higher scoring students and higher proportions of lower scoring students would bias the overall averages against the District. In order to get a fair estimate of District performance, comparisons to the State should be made strictly within subgroup.



Comparison to the State by Grade Level

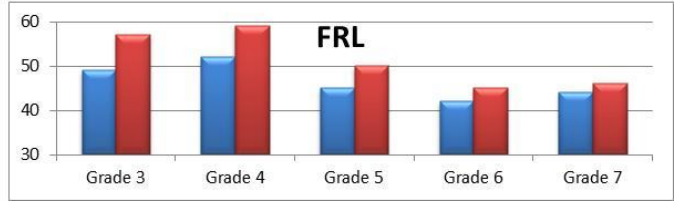
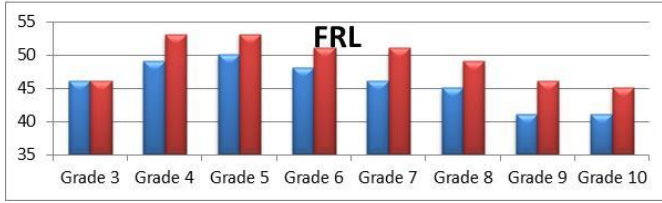
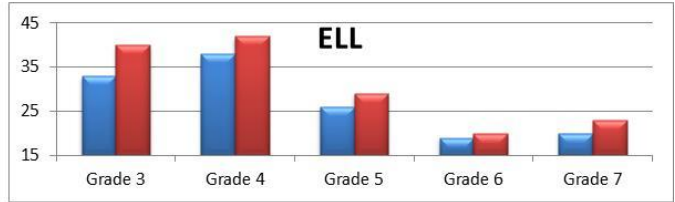
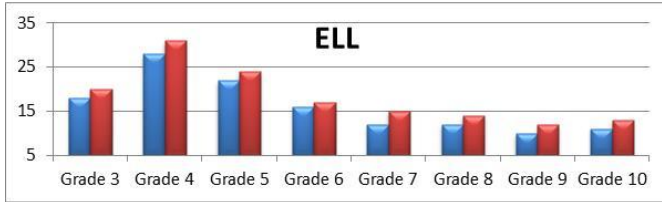
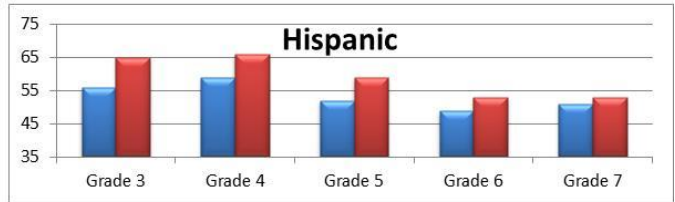
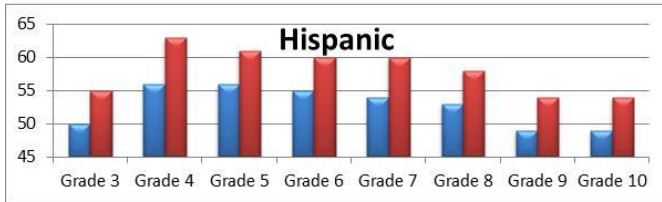
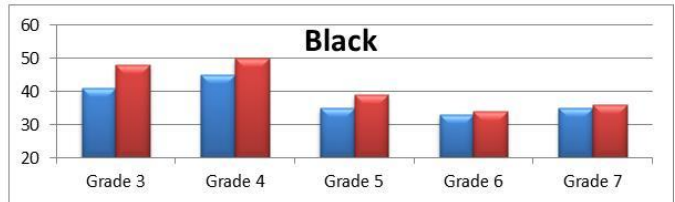
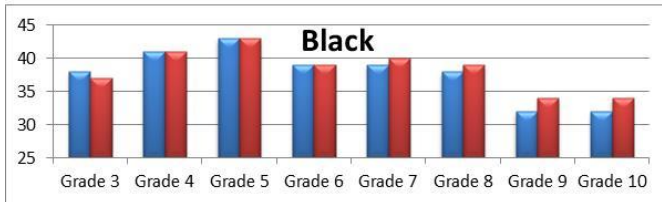
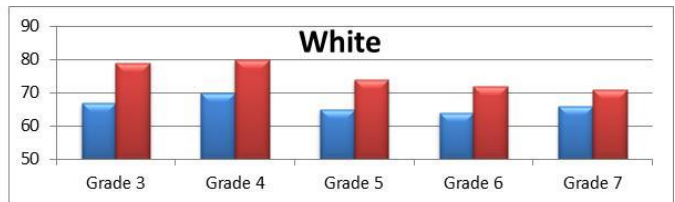
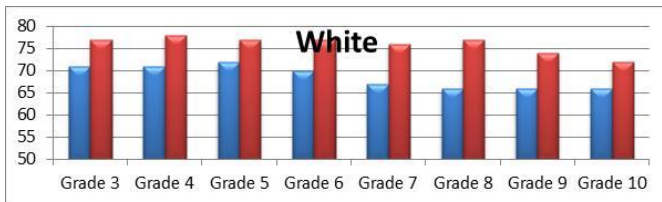
Because the performance levels of students differ from grade to grade, more precise comparisons can be made between the District and the State within each grade level. Below are graphs showing the percentages of students scoring at Level 3 or higher on the Reading and Mathematics FCAT in 2012-13. At all grade levels in Reading, the District looks worse than the State. In Math, the picture is slightly better, with the District outperforming the State in the lower grade levels. However, as we are now aware, these comparisons in which subgroups are aggregated are likely distortions of the real picture.



Below are the more appropriate comparisons within subgroup for Reading and Math for each grade level. It is overwhelmingly clear that the District outperforms the State in all subgroups. The exception, not shown here, would be in 8th grade for Math. That grade level is a special case. For the first time, in 2012-13 students had the option of taking *either* the FCAT 8th grade math test or the Algebra test. More of our higher performing students opted to take the Algebra test. In 2011-12, when all students who took the Algebra test also had to take the 8th grade Math test, the District also outperformed the State in all subgroups. Additionally, the same pattern holds true for Algebra, even though a substantially greater proportion of 8th graders in the District chose to take the Algebra test.

Reading
 ■ State ■ M-DCPS

Math
 ■ State ■ M-DCPS



Summary

Surprisingly, it is not uncommon for comparison trends in aggregate data to be in the reverse direction of trends within all component subgroups. The conditions for this to occur include disproportionate representation of differentially performing subgroups. This is exactly the conditions that hold when comparing the Miami-Dade School District to the State. While the District may appear lagging in global comparisons, we are, in fact, excelling in all categories when the data are broken down into customary subgroups.

The demographic makeup of our District is unique in the State. The educational demands of our District are particularly challenging. Evaluations made by combining all of our subpopulations into one big undifferentiated mass simply do not allow for the proper assessment of the progress of any constituent part of the District.

The discussion and the graphs presented in this paper make it clear that, in most cases, comparisons between the District and the State are appropriate only if they are made within subgroups. Extending this idea, it is advisable to confine appraisal to subgroups when comparing our district to other districts, or our city to other cities across the nation or in other countries. This advice applies not only to academic achievement data, but to safety records, health issues, resource management figures, and many other statistics of concern. Our efforts at guiding improvement require careful attention to each and all of the member groups that make up our District.