James P. Scanlan
Attorney at Law
1529 Wisconsin Avenue, NW
Washington, D.C. 20007
(202) 338-9224
jps@jpscanlan.com

September 12, 2016

Electronically Transmitted
Diane Gibson-Gray, President of the Board of Trustees
Stephanie Anello, Superintendent
Antioch Unified School District
510 G Street
Antioch, CA 94509

Re: Misunderstanding of Statistics Reflected in Actions of the East County NAACP Regarding Racial Differences in Discipline Outcomes in the Antioch Unified School District

Dear President Gibson-Gray and Superintendent Anello:

The principal purpose of this letter is to explain certain failures of understanding of elementary statistics reflected in actions taken by East County NAACP against Antioch Unified School District (AUSD) regarding AUSD discipline practices, including actions leading to an Interim Negotiated Settlement Agreement[1] (Agreement) executed between East County NAACP and AUSD on March 25, 2015, and the filing of a Complaint against AUSD by East County NAACP in Contra Costa Superior Court on July 6, 2016.

Like other actions taken against (and by) jurisdictions around the country regarding school discipline practices, the actions of East County NAACP regarding AUSD discipline practices are based on the belief, promoted by the United States Departments of Justice, Education, and Health and Human Services as well as numerous private institutions and organizations, that generally reducing student discipline rates will tend to (a) reduce relative (percentage) differences in discipline rates and (b) reduce the proportions groups most susceptible to adverse discipline outcomes comprise of persons experiencing those outcomes.

In fact, exactly the opposite is the case. Generally reducing discipline rates tends to increase relative differences in discipline rates and the proportions more susceptible groups comprise of persons disciplined.

[1] To facilitate consideration of issues raised in letters such as this by the addressees and others, I include links to referenced materials in electronic copies of the letters. Such copies may be found by means of the Measurement Letters page of jpscanlan.com. I generally also post links to such letters on the ASA Connect portion of the website of the American Statistical Association.

I have recently explained this matter fairly succinctly with regard to relative differences in adverse outcome rates of advantaged and disadvantaged groups in "Things government doesn't know about racial disparities," *The Hill* (Jan. 28, 2014), "The Paradox of Lowering Standards," *Baltimore Sun* (Aug. 5, 2013), and "Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies," *Amstat News* (Dec. 2012), and with regard to proportions disadvantaged groups make up of persons experiencing adverse outcomes in "Things the President Doesn't Know About Racial Disparities," *Federalist Society Blog* (Aug. 5, 2016), and "Things DoJ doesn't know about racial disparities in Ferguson," *The Hill* (Feb. 22, 2016). In explaining these issues, these items also describe the anomaly in federal civil rights law enforcement whereby, as a result of the government's failure to understand elementary statistics, entities that comply with government encouragements to relax standards and otherwise reduce the frequency of adverse outcomes in various settings increase the chances that the government (or others) will sue the entities for discrimination.

I have recently explained the pertinent, and related, statistical principles more elaborately in "The Mismeasure of Health Disparities," *Journal of Public Health Management and Practice* (July/Aug. 2016), "Race and Mortality Revisited," *Society* (July/Aug. 2014), "The Perverse Enforcement of Fair Lending Laws," *Mortgage Banking* (May 2014), "Measuring Health and Healthcare Disparities," Proceedings of the Federal Committee on Statistical Methodology 2013 Research Conference (March 2014) (FCSM paper), "The Mismeasure of Discrimination," Faculty Workshop, University of Kansas School of Law (Sept. 20, 2013) (Kansas Law paper), as well as in a letter to the American Statistical Association (Oct. 8, 2015)[2] and an amicus curiae brief in *Texas Department of Housing and Community Development, et al. v. The Inclusive Communities Project, Inc.*, Supreme Court No. 13-1731 (Nov. 18, 2016) (TDHCA brief). [3]

I have also recently explained the matter, by letter of July 3, 2016, to Jeffrey Sprague of the University of Oregon's Institute on Violence and Destructive Behavior, who is one of the experts identified in the Agreement and Complaint.[4]

---

[2] The American Statistical Association letter urges the organization, among other things, to explain to arms of the federal government that generally reducing adverse outcome rates tends to increase, not decrease, relative differences in adverse outcome rates and the proportions disadvantaged groups comprise of persons experiencing the outcomes. The letter has been referred to a subcommittee of the association's Scientific and Public Affairs Advisory Committee. A follow-up letter of July 25, 2016, urges the organization to explain the issue to President Barack Obama with regard to misunderstandings reflected in the President's statements on racial/ethnic differences in criminal justice outcomes.

[3] Older extended treatments of the issues may be found in "Can We Actually Measure Health Disparities?," *Chance* (Spring 2006), "Race and Mortality," *Society* (Jan./Feb. 2000), "Divining Difference," *Chance* (Fall 1994), "The Perils of Provocative Statistics," *Public Interest* (Winter 1991), and "The 'Feminization of Poverty' is Misunderstood," *Plain Dealer* (Nov 11, 1987).

[4] The letter to Professor Sprague was prompted by a June 2015 report of the University of Oregon Institute on Violence and Destructive Behavior and the University of Oregon Law School Center for Alternative Dispute Resolution titled "Eureka City Schools School-wide Positive and Restorative Discipline Assessment and

Section A of this letter discusses the principal measurement issues while explaining the widespread confusion in this area. Section B discusses the measurement issues in the context of racial differences in out-of-school suspensions in AUSD, with particular reference to data of the type discussed in Paragraphs 18 and 19 of the Complaint and the mistaken belief reflected in the former paragraph that generally reducing exclusionary discipline outcomes will tend to reduce the proportion African American students make up of persons experiencing those outcomes. In doing so, the section shows that, while the Complaint states that racial disparities in suspensions were essentially unchanged between the 2012-13 and 2014-15 school years, the lead expert identified in the Agreement and Complaint would have found substantial reductions in racial differences in suspension rates over that period. Section C discusses the measurement approaches of the discipline experts identified in the Agreement and Complaint and explains why these approaches will tend to reach opposite conclusions about whether racial differences are increasing or decreasing or about the comparative size of racial differences among different schools or different decision-makers or with respect to different types of conduct. Section D discusses the way that failure to understand the issues addressed in Section A generally undermines efforts to appraise racial differences in student outcomes apart from discipline.

A. Patterns by Which Measures of Differences in Discipline Outcomes Tend to Be Affected by the Frequency of Discipline Outcomes

All standard measures of differences between outcome rates (or between the proportion a group comprises of persons potentially experiencing an outcome the proportion it comprises of persons actually experiencing the outcome) tend to be systematically affected by the frequency of an outcome.

Most notably, inherent in other than highly irregular risk distributions is a pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. A corollary to this pattern is a pattern whereby the rarer an outcome the greater tend to be the proportions groups most susceptible to the outcome comprise of persons who experience the outcome and persons who avoid the outcome.

The patterns can be easily illustrated with normally distributed test score data. Table 1 below is based on a situation where an advantaged group (AG) and a disadvantaged group (DG) have mean test scores that differ by half a standard deviation (and where the two groups' distributions have the same standard deviation). At the higher cutoff the pass rates for AG and DG are 80% and 63% (with corresponding failure rates of 20% and 37%). Lowering the cutoff

---

Intervention Project Assessment Results and Programmatic Recommendations." The report, which was produced in connection with resolution of a suit regarding discipline issues against Eureka (CA) City Schools, reflected the mistaken view that the approaches to school discipline in the report's title, which tend generally to reduce overall discipline rates, will tend to reduce relative racial and other demographic differences in discipline rates.

to the point where 95% of AG passes would cause the pass rate of DG to rise to 87% (while the corresponding failure would be 5% and 13%). Column (5) shows that lowering the cutoff decreases the ratio of AG's pass rate to DG's pass rate from 1.27 to 1.09 (*i.e.*, reduces the relative difference in pass rates from 27% to 9%).[5] That lowering cutoffs tends to reduce relative differences in pass rates is a reason that lowering test cutoffs is universally regarded as reducing the disparate impact of tests on which some demographic groups outperform others. It may also be a reason that stringent standards are generally regarded as having an especially severe impact on disadvantaged groups.

But, as shown in column (6), lowering the cutoff increases the ratio of DG's failure rate to AG's failure rate from 1.85 to 2.60 (an increase in the relative difference in failure rates from 85% to 160%). Thus, lowering the cutoff reduced the relative difference in the outcome that increased in frequency (test passage) but increased the relative difference in the outcome that decreased in frequency (test failure).

Table 1. Illustration of effects on relative differences in pass and fail rates of lowering a cutoff from a point where 80% of AG passes to a point where 95% of AG passes, with proportions DG comprises of persons who pass and of persons who fail (when mean scores differ by approximately half a standard deviation and DG comprises 50% of test takers)

Cutoff	AG Pass (1)	DG Pass (2)	AG Fail (3)	DG Fail (4)	AG/DG Pass Ratio (5)	DG/AG Fail Ratio (6)	DG Prop of Test Takers (7)	DG Prop of Test Passers (8)	DG Prop of Test Failers (9)	Abs Diff (10)
High	80%	63%	20%	37%	1.27	1.85	50%	44%	65%	17
Low	95%	87%	5%	13%	1.09	2.60	50%	48%	72%	8

Columns 8 and 9 show how lowering the cutoff increases the proportion DG comprises of person who pass the test and the proportion DG comprises of persons who fail the test.[6] In circumstances where DG comprises 50% of the test takers, the proportion DG comprises of persons who pass the test would rise from 44% to 48%, while the proportion DG comprises of

[5] While I commonly refer to patterns of relative differences in this letter, the table actually presents rate ratios (also termed risk ratios). The relative difference is the rate ratio minus 1 where the rate ratio is above 1 and 1 minus the rate ratio where the rate ratio is below one. One should be careful not to mistakenly refer to the rate ratio as the relative difference. But the distinction between the two terms is not pertinent to the discussion here of patterns by which relative differences tend to be affected by the frequency of an outcome. In recent years I commonly present the rate ratios for both outcomes with the larger figure in the numerator, in which case, as to both outcomes, the larger the rate ratio, the larger the relative difference. Choice of numerator in the rate ratio, however, has no bearing on the patterns by which the two relative differences tend to be affected by the frequency of the outcome.

[6] Increasingly discipline disparities (as in Paragraphs 18 and 19 of the East County NAACP complaint) are analyzed in terms of differences between the proportion a group comprises of students and the proportion it comprises of students who are disciplines. See my letters to Department of Health and Human Services and Department of Education (Aug. 24, 2015) and Texas Appleseed (Apr. 7, 2015).

persons who fail the test increases from 65% to 72%. While it is necessary to posit the DG proportion of test takers to illustrate this point, the patterns of directions of changes in these proportions would hold irrespective of the actual proportion DG comprises of persons taking the test. And because lowering the cutoff does not change the proportion DG comprises of persons taking the test, lowering the cutoff will tend to reduce both the relative difference and the absolute difference between the proportion DG comprises of persons who take the test (the pool) and the proportion it comprises of persons who pass the test, while increasing both the relative difference and the absolute difference between the proportion DG comprises of the pool and the proportion it comprises of persons who fail the test.

Because the Agreement and Complaint identify Daniel Losen of the UCLA Center for Civil Rights and Remedies as the lead expert on discipline matters, and because Professor Losen commonly measures discipline disparities in terms of absolute (percentage point) differences between rates (as discussed further in Sections B and C), I have included in column 10 the absolute (percentage point) difference between rates. The size of the absolute difference is unaffected by which outcome one examines. But, like the other measures just discussed, the absolute differences tends to be systematically affected by the frequency of an outcome, though in a more complicated way than the other measures (as discussed, among other places, in each of the more extended references mentioned on page 2).

Roughly, viewing the matter in terms of an outcome that goes from being rare to being very common (while the opposite outcome changes in the opposite direction), absolute differences tend first to increase and then to decrease and tend also to track the direction of change of the larger relative difference (which is the relative difference associated with the less common of the corresponding outcomes). Since observers who rely on relative differences to measure disparities in most contexts other than testing, high school graduation, or healthcare,[7] tend to rely on the larger of the two relative differences, there is a tendency for observers relying on absolute differences to reach opposite conclusions about directions of change from persons relying on relative differences.

Graphical illustration of the relationship the absolute difference and the two relative differences (as well as the odds ratio[8]) according to the specifications underlying Table 1 above be found in Figure 3 (slide 36) of the January 2015 University of California, Irvine (UCI) methods workshop mentioned in note 9 *infra*.

[7] See "The Mismeasure of Health Disparities" and "Race and Mortality Revisited" regarding the shifting position of National Center for Health Statistics regarding whether to measure healthcare disparities in terms of relative differences in the favorable outcome (receipt of care) or relative differences in the adverse outcome (non-receipt of care).

[8] As the frequency of an outcome changes, the difference measured by the odds ratio tends to change in the opposite direction of the absolute difference. See note 22 *infra*.

For the rate ranges at issue in the example in Table 1 (which also are the rate ranges at issue in most analyses of discipline disparities) reducing the frequency of the less common outcome (which increases the frequency of the opposite, more common, outcome) will tend to reduce absolute differences between rates. In the hypothetical in Table 1, lowering the cutoff reduced the absolute difference from 17 to 8 percentage patterns.

The patterns shown with Table 1 are not peculiar to test score data or the numbers I chose to illustrate them. Rather, they exist in virtually every situation where two groups differ in their susceptibility to some favorable or corresponding adverse outcome, including school discipline outcomes.

Numerous tabular and graphical illustrations of the patterns, with actual or hypothetical data, may be found in the more extended references mentioned on page 2, in methods workshops I have given at American universities since 2012,[9] in the various pages and subpages of jpscanlan.com related to measurement issues, [10] and in many other letters to institutions or organization who activities, or whose members' activities, involve or are affected by analyses of demographic differences.[11] [12] Among letters of particular pertinence to the instant situation are

[9] See "The Mismeasure of Health Disparities in Massachusetts and Less Affluent Places," Department of Quantitative Health Sciences, University of Massachusetts Medical School (Nov. 18, 2015); "The Mismeasure of Discrimination," Center for Demographic and Social Analysis, University of California, Irvine (Jan. 20, 2015); "The Mismeasure of Demographic Differences in Outcome Rates" Public Sociology Association of George Mason University (Oct. 18, 2014); "Rethinking the Measurement of Demographic Differences in Outcome Rates," Maryland Population Research Center of the University of Maryland (Oct. 10, 2014); "The Mismeasure of Association: The Unsoundness of the Rate Ratio and Other Measures That Are Affected by the Prevalence of an Outcome," Minnesota Population Center and Division of Epidemiology and Community Health of the School of Public Health of the University of Minnesota (Sept. 5, 2014); "The Mismeasure of Group Differences in the Law and the Social and Medical Sciences," Institute for Quantitative Social Science at Harvard University (Oct. 17, 2012); "The Mismeasure of Group Differences in the Law and the Social and Medical Sciences," Department of Mathematics and Statistics of American University (Sept. 25, 2012).

[10] The principal measurement pages are: Measuring Health Disparities, Scanlan's Rule, Mortality and Survival, Statistical Reasoning, Immunization Disparities, Educational Disparities, Disparate Impact, Discipline Disparities, Lending Disparities, Employment Discrimination, Feminization of Poverty, and Vignettes. The pages have close to a hundred subpages.

[11] Recipients of such letters include American Statistical Association II (July 25, 2016), Federal Judicial Center (July 7, 2016), University of Oregon Institute on Violence and Destructive Behavior and University of Oregon Law School Center for Dispute Resolution II (July 5, 2016), University of Oregon Institute on Violence and Destructive Behavior and University of Oregon Law School Center for Dispute Resolution (July 3, 2016), New York City Center for Innovation through Data Intelligence (June 6, 2016), Consortium of Social Science Associations (Apr. 6, 2016), Population Association of America and Association of Population Centers (Mar. 29, 2016), Council of Economic Advisers (Mar. 16, 2016), City of Madison, Wisconsin (Mar. 12, 2016), Stanford Center on Poverty and Inequality (Mar. 8, 2016), City of Boulder, Colorado (Mar. 5, 2016), Houston Independent School District (Jan. 5, 2016), Boston Lawyers' Committee for Civil Rights and Economic Justice (Nov. 12, 2015), House Judiciary Committee (Oct. 19, 2015), American Statistical Association (Oct. 8, 2015), Chief Data Scientist of White House OSTP (Sept. 8, 2015), McKinney, Texas Independent School District (Aug. 31, 2015), Department of Health and Human Services and Department of Education (Aug. 24, 2015), Agency for Healthcare Research and Quality (July

each of those to individual school districts or boards of education, as well the above-mentioned July 3, 2016 letter to Professor Sprague (identified in note 11 as that July 3, 2016 letter to the University of Oregon Institute on Violence and Destructive Behavior and University of Oregon Law School Center for Dispute Resolution), and the letters to Boston Lawyers' Committee for Civil Rights and Economic Justice (Nov. 12, 2015), Department of Health and Human Services and Department of Education (Aug. 24, 2015), and Texas Appleseed (Apr. 7, 2015). All of the letters in note 11, however, illustrate how universally analyses of demographic differences involving favorable or adverse outcomes are undermined by failure to understand patterns by which measures tend to be affected by the frequency of an outcome.

Examples of these patterns based on actual data that are of particular pertinence to the instant situation may be found in the subpages of the Discipline Disparities page of jpscanlan.com discussing the way that recent reductions in discipline rates around the country have been accompanied by increased relative differences in discipline rates: California Disparities, Colorado Disparities, Connecticut Disparities, Maryland Disparities, Minnesota Disparities, Oregon Disparities. Beaverton, OR Disparities, Denver Disparities, Henrico County, VA Disparities, Los Angeles SWPBS, Minneapolis Disparities, Montgomery County, MD Disparities, Portland, OR Disparities, St. Paul Disparities.[13] While these pages discuss the matter in terms of relative differences in discipline rates, for reasons discussed above, increased relative differences in discipline rates correlate with increases in the proportions more susceptible groups

---

1, 2015), City of Minneapolis, Minnesota (June 8, 2015), Texas Appleseed (Apr. 7, 2015), Senate Committee on Health, Education, Labor and Pensions (Mar. 20, 2015), United States Department of Justice and City of Ferguson, Missouri (Mar. 9, 2015), Vermont Senate Committee on Education (Feb. 26, 2015), Portland, Oregon Board of Education (Feb. 25, 2015), Wisconsin Council on Families and Children's Race to Equity Project (Dec. 23, 2014), Financial Markets and Community Investment Program, Government Accountability Office (Sept. 9, 2014), Education Law Center (Aug. 14, 2014), IDEA Data Center (Aug. 11, 2014), Institute of Medicine II (May 28, 2014), Annie E. Casey Foundation (May 13, 2014), Education Trust (April 30, 2014), Investigations and Oversight Subcommittee of House Finance Committee (Dec. 4, 2013), Mailman School of Public Health of Columbia University (May 24, 2013), Senate Committee on Health, Education, Labor and Pensions (Apr. 1, 2013), Federal Reserve Board (March 4, 2013), Harvard University et al. (Oct. 26, 2012), Harvard University (Oct. 9, 2012), United States Department of Justice (Apr. 23, 2012), United States Department of Education (Apr. 18, 2012), The Commonwealth Fund (June 1, 2010), Institute of Medicine (June 1, 2010), National Quality Forum (Oct. 22, 2009), Robert Wood Johnson Foundation (Apr. 8, 2009).

[12] I have been led to understand that persons with formal statistical training may best under the described pattern by which relative differences tend to be affected by the frequency of an outcome in terms of the density function. Such is the approach of recent papers co-authored by University of Oregon Professor of Economics Peter J. Lambert. See Lambert PJ, Subramanian S (Disparities in Socio-Economic outcomes: Some positive propositions and their normative implications. Soc Choice Welf 2014;43:565-576), and Lambert PJ, Subramanian S (Group inequalities and "Scanlan's Rule": Two apparent conundrums and how we might address them. Working Paper 84/2014, Madras School of Economics (2014)).

[13] Reportage of situations where general reductions in discipline rates have been accompanied by reduced racial differences in discipline have generally involved the work by the UCLA Center for Rights and Remedies (of which Daniel Losen is Director), which commonly measures disparities in terms of absolute differences between rates.

comprise of persons disciplined. See also the DOE Equity Report subpage regarding a Department of Education report showing that relative racial differences in expulsions are larger in school districts without zero tolerance policies than school districts with such policies; the Suburban Disparities subpage regarding the greater relative racial differences in suspensions in suburban than central city schools; "Race and Mortality Revisited" (at 342) regarding the larger relative differences in suspension rates in pre-school than in K-12; and the Boston Lawyers' Committee letter regarding larger relative differences in suspensions in Massachusetts (which has generally low suspension rates) than nationally.

The described patterns by which measures tend to be affected by the frequency of an outcome, of course, will not always be observed when there occur changes in the frequency of an outcome (or in comparisons of settings with different frequencies of an outcome.) For the strength of the forces causing the outcome rates to differ may change over time and otherwise vary from setting to setting (as shown in many example in materials referenced above and as I will show with AUSD data in Section B). But data on demographic differences can only be soundly analyzed while understanding these patterns and their implications.

The most effective way to quantify the strength of the forces causing outcome rates to differ (which might also be characterized as the difference in the circumstances of two groups reflected by their outcome rates) is by deriving from the adverse (or corresponding favorable) outcome rates of the advantaged and disadvantaged groups the difference, in terms of percentage of a standard deviation, between means of the hypothesized underlying distributions. This method is commonly referred to as the probit and I generally describe the results as the "EES" for "estimated effect size." The method is used with a wide variety of data in the most of the materials referenced in the text above (apart from the short items mentioned at the top of page 2). It will be used with regard to AUSD discipline data in Section B.[14]

Implicit or explicit in the above discussion and the materials it references are the following things that must be understood in interpreting data on demographic differences in discipline rates. First, to the extent that the forces causing the outcome rates of the advantaged and disadvantaged groups in Table 1 to differ (whether the rates involve test outcomes or anything else) can be measured, there is no basis to distinguish between those forces from row to row of the table.

Thus, for example, if the failure rates in the Table 1 reflect suspension rates at two points in time, there would be no basis for arguing that the strength of the forces causing the outcome rates to differ (whatever those forces may be) had changed over time or any basis for drawing inferences about processes based on the changes over time. The same holds for comparisons of

[14] I do not present an EES figure in Table 1 because, since the outcome rates are derived from a situation where the mean test scores of AG and DG differ by half a standard deviation, the EES would necessarily be .5. An EES value of .5 corresponds to a situation where approximately 31 percent of DG falls above the mean for AG. See Table 7 (slide 47) of the UCI workshop for illustrations of the meanings of various EES values in terms of the proportion of DG falling above the AG mean.

racial differences in outcome rates at different schools, by different decision-makers, or with respect to different types of behavior. By way of a simple example, other things being equal, the teacher who is a strict disciplinarian will tend to show patterns of outcome rates and measures thereof more like those in first row of the table while the teacher who is less strict a disciplinarian will tend to show patterns more like those in the second row.

Understanding the point about particular teachers is potentially of substantial importance with regard to AUSD dealings with East County NAACP given the apparent concern on the part of AUSD (discussed at pages 13-14 of the Complaint) that teachers identified as making discriminatory discipline decisions would be subject to disciplinary action. See the discussion in "Race and Mortality Revisited" regarding its Table 5 and the Kansas Law paper regarding its Table 1 regarding the way that situations where discrimination at issue best illustrate the fatal flaws of standard measures of differences between outcome rates and the absurdity of notions that measures that yield opposite conclusions about the comparative size of disparities might all be in some respect valid.

Second, as discussed above and in the longer references, in order to estimate the strength of the forces causing outcome rates to differ one must have the outcome rates themselves. Information on the proportion a group comprises of persons potentially experiencing an outcome and the proportion it comprises of persons actually experiencing an outcome do not enable one to derive these rates. Thus, such information can never be effectively analyzed even though one may be able to observe certain expected patterns in such data. See Section of TDHCA brief (at 23-27) and Section C the Kansas Law paper (at 23-26). Further, there are additional problems in analyses based on those proportions apart from that just mentioned, including, among others, that the same pair of outcome rates for an advantaged and disadvantaged group will yield different conclusions about the size of the disparities depending on the proportion the disadvantaged group makes up of the pool. See the IDEA Data Center Disproportionality Guide and subpage of the Discipline Disparities page of jpscanlan.com and slides 98-108 of the University of Maryland Workshop mentioned in note 10.[15] Even apart from these additional problems, however, one should never attempt to appraise the size of demographic differences based on a measure of the difference between the proportion a group comprises of the pool and the proportion the group comprises of persons experiencing an outcome.

B. Illustrations Based on AUSD Suspension Data

Paragraph 18 of the East County NAACP Complaint lists the "use of exclusionary discipline rather than educational interventions to address behavioral issues" as among factors

[15] Here, too, focus on a situation where discrimination is at issue usefully illustrates invalidity of certain measures. For appraisals of the likelihood of discrimination based on comparison of the proportion a group comprises of the pool and the proportion it comprises of persons experiencing an outcome involving any given pair of outcome rates for the advantaged and disadvantaged groups would vary depending on the proportion the disadvantaged group comprises of the pool.

that harm African American students by producing disparate rates of suspension and expulsion. Paragraphs 18 and 19 then cite, among other data, the proportions African Americans comprise of students in the 2012-13 and 2014-15 school years and the proportions African Americans made up of all students receiving out-of-school suspensions in those years, describing the patterns as essentially unchanged over the period covered.[16]

The figures from Paragraphs 18 and 19 are included in columns 7 and 9 of the first and third rows of Table 2, which tracks the format of Table 1.[17] Table 2 also includes the actual suspension rates for the referenced school years, as well as 2013-14, along with the other information presented in Table 1, as well as the EES figure explained above.

Table 2. White and African American rates of receiving no suspensions and receiving at least one suspension in the school years 2012-13, 2013-14, and 2015, with measures of difference

Period	W No Susp (1)	AA No Susp (2)	W Susp (3)	AA Susp (4)	W/AA NoSusp Ratio	AA/W NoSusp Ratio	AA Prop of Students (7)	AA Prop of Non-Susp (8)	AA Prop of Susp (9)	Abs Diff (10)	EES (11)
12-13	89.3%	59.3%	10.7%	40.7%	1.50	3.78	24.8%	17.9%	57.3%	29.9	1.00
13-14	92.4%	66.8%	7.6%	33.2%	1.38	4.37	25.3%	19.7%	60.4%	25.5	1.00
14-15	92.2%	70.6%	7.8%	29.4%	1.31	3.76	26.0%	21.0%	60.0%	21.6	0.89

Between the 2012-13 and 2013-14 school years, there occurred substantial declines in suspension rates. During that period there occurred the usual pattern of changes in measures of difference in the circumstances. Relative differences in the increasing outcome (avoiding suspension) decreased while relative difference in the decreasing outcome (suspension) increased (as shown in columns (5) and (6)). And the proportions the group more susceptible to the

[16] At least one reader of Paragraphs 18 and 19 confused the proportion African Americans make up of students suspended with the African American suspension rate. Alluding to Paragraphs 18 and 19 of the Complaint, an *EdSource* article titled "NAACP lawsuit alleges that Antioch Unified violated school discipline agreement" stated that, "while the total number of students suspended or expelled in the district has dropped in the last few years, the rate of suspensions for African- American students has remained essentially unchanged, the lawsuit said." Actually the Complaint merely stated that the proportion African Americans made up of suspended students had remained essentially unchanged. As shown in Table 2, the African American suspension rate declined from 40.7% to 29.4%. See the July 1, 2016 Prefatory Note to the Rhode Island Disparities subpage of the Discipline Disparities page regarding a situation where similar confusion in a study by the Rhode Island NAACP made a commonplace pattern appear to be an extraordinary pattern.

[17] Presentation of the non-suspension rates (and relative differences between those rates) before the information on suspension rates may seem unusual to some readers. But I believe will be easier to understand if I follow the approach of Table 1. That approach is a function of my preference for explaining effect of lowering a cutoff on relative differences in pass rates (something that is commonly understood) before the effect on relative differences in failure rates (something that almost no one understands).

decreasing outcome comprised of persons avoiding the outcome and experiencing the outcome both increased (as shown in Columns (9) and (10)).[18] These patterns would tend to occur regardless of the rate ranges at issue.

The table also shows the common pattern of changes in the absolute difference for the rate ranges at issue, which is a decrease in such difference (as shown in Column 10). The EES (1.00 standard deviations in both years) indicates that to the extent that the strength of the forces causing the rates to differ can be quantified, it changed not all between the two school years.

The table tells a somewhat different story regarding changes between the 2013-14 and 2014-15 school years, when discipline rates continued to decline. In this case both the relative difference in the favorable outcome and the relative difference in the adverse outcome decreased (as did the absolute difference), a pattern suggesting a genuine decrease in the strength of the forces causing the outcome rates of African American and white students to differ. To the extent that the strength of the forces can be quantified, it declined from 1.00 standard deviations to .89 standard deviations.

Notably, over the two school year period discussed in Paragraphs 18 and 19 of the Complaint, the absolute difference between the African American and white rates decreased from 29.9 to 21.6 percentage points. Thus, Mr. Losen, the lead discipline expert identified in the Complaint, would regard the disparity to have decline substantially while the Complaint regards the disparity to be essentially unchanged.[19]

[18] Analyses of the difference between the proportion African Americans comprise of the pool and the proportion they comprise of persons experiencing the favorable and adverse outcomes is complicated by the fact that, in contrast to the hypothetical reflected in Table 1, the proportion the disadvantaged group comprised of persons in the pool increased slightly between the two school years, a factor that, irrespective of other matters, would tend to increase the proportion the disadvantaged group makes up of persons experiencing both outcomes. (It also warrants note that Table 2 differs from Table 1 in that Table 1 involved a universe comprised of only two groups while the data on proportions in Table 2, which I use for consistency with Paragraphs 18 and 19, involve several groups.) But it does appear that both the relative and absolute difference between the proportion African Americans comprised of the pool and the proportion they comprised of persons experiencing the favorable outcome decreased (from 38.9% to 28.8% for the former and from 6.9 to 5.7 percentage points for the latter), while those differences for the adverse outcome increased (from 31% to 38% for the former and from 32.5 to 35.0 percentage points for the latter). But given the complexity of these measures and the fact that (as discussed toward the end of the prior section) such measures are especially problematic, I do not treat these measures in the body of this letter.

[19] While a conclusion that the disparity declined would be correct in the circumstances, the absolute differences does not reflect the extent of the change. For the 8.3 percentage point reduction (which might be characterized as a 28% reduction) is a function of both the change in the frequency of an outcome and the change in the strength of the forces causing the outcome rates to differ.

C. The Measurement Approaches of the Disparities Experts Identified in the Agreement and Complaint

As reflected in the discussion in the more extensive references mentioned on page 2 and as reflected in the letters listed in note 11, even where analyses of demographic differences do not involve beliefs about the effects of the frequency of an outcome on particular measures that are the opposite of reality, the research on demographic differences involving outcome rates has generally provided little of value and much that is misleading about the size disparities, how they may changing over time, or how they may be vary from setting to setting. For persons analyzing demographic differences in outcomes almost invariably do so while not understanding patterns by which measures tend to be affected by the frequency of an outcome and while never exploring the extent to which an observed pattern is a function of the frequency of an outcome and the extent to which it may indicate something meaningful about underlying processes.

But there are other reasons as well why the study of demographic differences has confused observers (including observers who do not know they are confused). Researchers commonly report results based on their preferred measure as if it were the only possible measure, without mention that other measures may – or in fact do in the particular study – yield opposite conclusions, even when the other measures are more commonly used in the circumstances. Researchers also commonly cite patterns in studies by others without any consideration of the measures employed in those studies study and whether such measures tend to (or in fact did) yield opposite conclusions from the measure employed by the researchers citing the studies. Illustrations of the degree of may be found on the AHRQ's Vanderbilt Study of the Measuring Health Disparities page and the Spurious Contradictions subpage of the Scanlan's Rule page of jpscanlan.com. See also the letter to Agency for Healthcare Research and Quality (July 1, 2015) regarding a situation where confusion about relative and absolute differences caused the National Healthcare Disparities Report to highlight as some of the largest reductions in healthcare disparities over a particular period situations where the report would also regard the disparity to be much larger at the end of the period than at the beginning of the period.

As discussed in "The Mismeasure of Health Disparities," "Race and Mortality Revisited," and the FCSM paper, lately health and healthcare disparities research have been giving increasing attention to situations where one would reach different conclusions about directions of changes in disparities depending on whether one relies on the absolute difference or the relative difference the observer happens to be examining (commonly the larger relative difference). They have done so, however, without acknowledging a second relative difference. They have also shown no evident awareness that, while both of the relative differences and the absolute difference may all change in the same direction (as in the case of the patterns reflected in the second and third rows of Table 2), anytime the mentioned relative difference and the absolute difference have changed in a different directions, the unmentioned relative difference will necessarily have changed in the opposite direction of the mentioned relative difference and the same direction as the absolute difference.

Studies of discipline disparities (and some other educational disparities, as discussed in Section D) remain an area where there is extremely limited recognition of the possibility (or actuality) that relative differences in the adverse outcomes usually examined and absolute differences between rates will commonly yield opposite conclusions. Indeed, abstracts typically do not even state what measure was used. Those understanding the patterns described here can generally infer that studies reporting greater racial disparities in suspensions in lower grades than higher grades or in suburban schools than central city schools or increasing disparities during periods of decreasing adverse outcomes have relied on relied on relative differences in adverse outcomes (or differences between the proportions disadvantaged groups comprise of students and the proportions they comprises of persons experiencing adverse outcomes),[20] while studies reporting opposite patterns have relied on absolute differences between rates.

As mentioned already, Daniel Losen of the UCLA Center for Civil Rights and Remedies UCLA , who is identified as the lead discipline expert in the Agreement and Complaint, relies principally on absolute differences between rates.[21] The report Dr. Losen authored on the Syracuse (NY) City School District ("Getting Back on Track, the Syracuse Report on School Discipline"), which Section 2.3 of the agreement suggests as a model for the report Mr. Losen is to produce on the discipline practices of AUSD, relies on absolute differences between rates. Thus, as already suggested, Mr. Losen will tend to reach opposite conclusions about the comparative size of disparities from persons relying on relative differences in discipline rates (or the corresponding patterns of differences between the proportion a group comprises of the pool and the proportion it comprises of persons experiencing an adverse discipline outcome).

---

[20] I include the parenthetical because an important point of this letter is the tendency for the measurement approach in the Complaint to reach opposite conclusions from those reached by Mr. Losen. But the inclusion somewhat oversimplifies certain matters. Suburban schools, which tend to have lower overall suspension rates than central city schools, will tend to show larger relative racial/ethnic differences in suspension rates than central city schools, as discussed on the Suburban Disparities subpage of the Discipline Disparities page. The larger relative difference in suburban schools will tend toward causing racial/ethnic minorities to comprise of a higher proportion of persons suspended in suburban than central city schools. But the higher minority representation in central city schools than in suburban schools will tend to reduce the relative difference between the proportion racial minorities make up of the pool and the proportion they comprise of persons suspended and have varying effects on the absolute differences between those proportions. See Table 2 of IDEA Data Center Disproportionality Guide subpage of the Discipline Disparities page.

[21] The NEPC National Study subpage of the Discipline Disparities page (originally posted in May 2012) discusses a 2011 study by Mr. Losen that found general increases in suspension rates since the early 1970s to increase racial disparities in discipline rates, which Mr. Losen measured in terms of absolute differences between rates. Just as decreases in discipline rates will tend to reduce absolute differences, increases in discipline rates will tend to increase absolute differences. As discussed on that subpage, however, data in Mr. Losen's study also showed increases in relative differences in discipline rates, which is a departure from frequency-related patterns described earlier. The subpage posits some possible explanations for that departure.

In contrast to most persons analyzing demographic differences in discipline rates, who commonly ignore all measures other than the one they happen to be employing, Mr. Losen has discussed the potential for different interpretations as to directions of changes in discipline disparities over time depending on whether one examines relative differences in adverse outcomes or the absolute difference. A February 2015 study titled "Are we closing the school discipline gap," authored by Mr. Losen and colleagues at the UCLA Center for Civil Rights and Remedies, analyzed changes in discipline disparities in terms of absolute differences between rates (though the study (at 6) makes a reference to disparities in suspensions between students with and without disabilities in terms of the ratio of the former to the latter, a ratio that general decreases in suspensions will tend to increase).

But the study's Appendix B discusses (at 48) the possibility for relative and absolute differences to yield opposite conclusions as to directions of changes, positing a situation where initially the black and white suspension rates are 40% and 20% and then drop to 3% and 1% respectively. In this hypothetical the absolute difference decreased from 20 to 2 percentage points while the risk ratio increased from 2.0 to 3.0. The study, which describes the disparity as reduced to one tenth of its original size, indicates that it favors the absolute difference as a measure of disparity in these terms: "While we agree that a risk ratio of 3.0 does suggest that a problem remains, we assert that progress has been made when suspension rates go down and the racial gap [in absolute terms] narrows."

The discussion does not reflect an understanding of the ways the absolute difference tends to change in the opposite direction of the relative difference in suspension rates or of the pattern by which the absolute difference tends to be affected by the frequency of an outcome and the implications of such pattern with regard to the utility of the absolute difference as a measure of association. As it happens, the situation posited would involve a reduction in disparity (EES reduced from .59 to .45).[22] But if the black rate declined from 40% to 25% while the white rates declined from 20% to 7% – a reduction in the absolute difference from 20 to 18 percentage points – the EES would have increased from .59 to .80.

The last example illustrates one of the reasons it is why it is impossible to effectively evaluate changes in forces causing outcome rates to differ without understanding the patterns

[22] Because of its complexity, I commonly give only limited attention to the odds ratio, though some would maintain that it provides a satisfactory alternative to the probit. As discussed in many of the longer references and note 8 *supra* (and as shown in the aforementioned Figure 3 of the UIC workshop), as the prevalence of an outcome changes, differences measured by the odds ratio tend to change in the opposite directions of the absolute difference. When neither outcome is uncommon, the difference measured by the odds ratio gives results similar to the probit. But when either outcome is uncommon, the difference measured by the odds ratio tends to behave like the larger relative difference. In Mr. Losen's hypothetical, the described reductions in discipline rates would cause the odds ratio to increase from 2.67 to 3.06. In my view, results yielded by odds ratio are better presented in terms of the standardized mean difference that can be derived from the odds ratio. In the Losen hypothetical, with the higher suspension rates, the standardized mean difference (at .54 standard deviations) does not differ greatly from the .59 EES. But, whereas the hypothesized reduction in suspension rates caused the EES to decrease from .59 to .45, it caused the standardized mean difference to increase from .54 to .65.

described in the references discussed above. "Race and Mortality Revisited shows in its Table 4 (in the top two rows) and Table 6 situations where absolute differences increased even though there occurred a decrease in the strength of the forces causing the outcome rates to differ.[23] Problems arising the from the failure to understand patterns by which measures tend to be affected by the frequency of an outcome include (in addition to never effectively quantifying the size of disparities or changes therein) that one may not only identify changes in disparities that do not occur and fail to identify changes that do occur, but may also identify changes in one direction when the change actually is in the opposite direction.

The key point of the instant section, however, is that the approach to measuring to measuring discipline disparities of the lead expert identified in the Agreement and Complaint will tend, not only to reach different conclusions from that yielded by the approach reflected in the Complaint (as discussed in Section b), but to reach different conclusions from the other experts identified in the Agreement and Complaint.

As indicated in the above-mentioned letter to Professor Sprague, he and his colleagues at the University of Oregon's Institute on Violence and Destructive Behavior appear to employ the common approach of measuring discipline disparities in terms relative differences in discipline rates. Thus, in attempting to determine whether disparities are increasing or decreasing over time, or are otherwise larger in one setting than another, and whether particular policies tend to be associated with larger or smaller disparities, Professor Sprague will tend to reach opposite conclusions from Mr. Losen.

I am not familiar with work of persons identified as experts from the Berkeley Haas Institute for a Fair and Inclusive Society and the Center for Policing Equity or the Research for Perception Institute. But I am generally familiar with any work that recognizes the implications of patterns by which measures tend to be affected by the frequency of an outcome – or, most germane to the instant matter, recognizes that reducing adverse outcomes tends to increase, not decrease, relative differences in rates of experiencing the outcome and the proportion disadvantaged groups make of persons experiencing the outcomes. And I am not aware that any of the identified experts from the referenced organizations understands these issues.

I note that an article titled "Implicit Bias Insights as Preconditions to Structural Change," by john powell of the Berkeley Haas Institute for a Fair and Inclusive Society makes this statement: "Black and Latino men and women serve as executives at Fortune 500 companies and as presidents of our finest universities, yet Black and Latino children are 3 times as likely to live in poverty and 20% less likely to graduate from high school than White children." Thus, Professors powell and Godsil appear in this instance to discuss one issue in terms of relative differences in the adverse outcome and the other in terms of relative differences in the favorable

[23] "Race and Mortality Revisited" (at 337-339) also discusses the way that reliance on absolute differences to measure healthcare disparities has caused Massachusetts to include a disparities provision in its Medicare pay-for-performance program that will tend to increase healthcare disparities,

outcome.[24] Presumably, in the case of discipline disparities, Professors powell and Godsil (and the other expert identified in Section 1.c of the agreement as affiliated with the Haas Institute) will rely on relative differences in adverse discipline outcomes with respect to appraisal of the matters mentioned at the end of the prior paragraph. They will thus tend to reach conclusion that are the same as those Professor Sprague would reach and the opposite of those Mr. Losen would reach with regard to those matters.

Like the overwhelming majority of other experts in the analyses of demographic differences, however, none of the experts identified in the Agreement or Complaint has indicated an awareness of the patterns by which measures tend to be affected by the frequency of an outcome or how one might quantify demographic differences while taking those patterns into account.

D. Measurement Issues Regarding Demographic Differences in Educational Outcomes Apart From Discipline

The discussion in Section A would apply generally to analyses of disproportionality is special education, which consume substantial resources. But schools also give substantial attention to racial/ethnic differences in academic outcomes, sometimes appraising these outcomes in terms of relative differences in favorable outcome and sometimes in terms of relative differences in adverse outcomes. An item of particular pertinence in this area is the letter to New York City Center for Innovation through Data Intelligence (June 6, 2016), which discusses a recent study of racial/ethnic difference in educational outcomes among young people in New York City that sometimes measured disparities in terms of relative differences in favorable outcomes like graduation and sometimes measured closely related matters in terms of relative difference in adverse outcomes like dropping out of school.. The study, which found that educational outcomes were generally improving (*i.e.*, increases in favorable outcomes and decreases in the corresponding adverse outcomes), tended to find reductions in disparities it measured in terms of relative differences in favorable outcomes and increases in disparities it measured in terms of relative difference in adverse outcomes. But, as a result of the failure to recognize patterns by which measures tend to change as the frequency an outcome changes, the study was not able actually to provide useful information of whether educational differences between advantaged and disadvantaged groups reflected by the groups' outcome rates were increasing or decreasing. See also the Harvard CRP NCLB Study and the McKinsey Achievement Gap Study subpages of the Educational Disparities page of jpscanlan.com regarding situations where observers relied on relative differences in favorable or adverse

[24] I note that income data provides one of the best means of showing how reducing the frequency of an outcome tends to increase relative differences in rates of experiencing the outcome while reducing relative differences in rates of avoiding the outcome (and increases the proportion disadvantaged groups comprise of persons experiencing and avoiding the outcome). I may use such data more often than test score data. See, e.g., Table 2 of "Race and Mortality Revisited," Table 1 of "Can We Actually Measure Health Disparities?," and Figure 7 (slide 63) of the University of Maryland workshop identified in note 10.

outcomes to appraise demographic differences in academic outcomes without recognizing the effect of the frequency of the outcome on the measure employed.

The most common approach to measuring disparities in rates of meeting or failing to meet various proficiency standards involves reliance on absolute differences between rates. Almost invariably, those relying on absolute differences do so without recognizing that improvements in favorable outcome rates that are very low will tend to increase absolute differences while improvements in favorable outcome rates that are generally high will tend to reduce absolute differences. There seems no recognition, for example, that general improvements in educational outcomes will tend to increase absolute differences in rates of meeting advanced levels of proficiency (where rates tend to be quite low) but reduce absolute differences in rates of meeting the basic proficiency level (where rates can often be well above 50% for all groups). Such issues are discussed on the main Educational Disparities page and its Disparities by Subject, New York Proficiency Rate Disparities, Education Trust High Achiever Study , Education Trust Glass Ceiling Study, Education Trust High Achiever Study, Education Trust Glass Ceiling Study, Annie E. Casey 2014 Proficiency Disparities Study subpages, usually with regard to the fact that studies that fail to understand the patterns by which absolute differences tend to be affected by the frequency of an outcome are unable to provide useful information on whether differences in the educational circumstances of advantaged and disadvantaged groups are increasing or decreasing over time or are larger in one setting than another. See also the letters to the Annie E. Casey Foundation (May 13, 2014) and Education Trust (April 30, 2014). But see pages 3-4 of the letter to Stanford Center on Poverty and Inequality (Mar. 8, 2016) regarding recent work of Professors of Education at Harvard and Stanford who have recognized the way absolute differences between proficiency rates will tend to be affected by the general proficiency level and have suggested approaches to measurement akin to the approach that yields the EES figure discussed above.

Materials discussed above in this section generally involve appraisal of academic (non-discipline) disparities based on outcomes rates. But sometime observers will discuss such disparities in terms of comparison between the proportion a group makes up of the pool and the proportion persons make up of persons experiencing an outcome, and such comparisons can be made with regard to both favorable outcomes and adverse outcomes. For example, it might be noted that a group that comprises 20% of students comprises only 10% of persons taking Advanced Placement courses or that the same group comprises 40% of persons retained in grade. Thus, it is important to understand that general improvements in education that increase the proportion of students experiencing favorable outcomes and decrease the proportion of student experiencing adverse outcomes will tend to increase the proportion disadvantaged groups make up of persons experiencing the outcomes described in favorable terms (hence reducing all measures of differences between the proportion the group comprises of the pool and the proportion it comprises of persons experiencing such outcomes) but increase the proportion such groups comprise of persons experiencing outcomes described in adverse terms (hence increasing all measures of differences between the proportion the group comprises of the pool and the

proportion it comprises of persons experiencing such outcomes). And when outcomes generally worsen rather than improve, one will commonly observe the opposite patterns.

While I stress the importance of understanding these patterns of disproportionality, I note again the point made at the end of Section A that one ought never to analyze demographic differences while focusing on the proportion a group comprises of the pool and the proportion it comprises of persons experiencing either outcome, rather than on the rates at which the groups experience either outcome.

Sincerely,

/s/ James P. Scanlan

James P. Scanlan

.