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

January 5, 2016

ELECTRONICALLY TRANSMITTED
Rhonda Skillern-Jones, President
Board of Trustees
Houston Independent School District
Hattie Mae White Ed. Support Ctr.
4400 West 18th Street
Houston, Texas 77092-8501
Rskille2@houstonisd.org

Re: Misunderstanding of Statistics in December 7, 2015 Letter from Texas Appleseed to Houston Independent School District

Dear President Skillern-Jones:

This letter has two purposes. One purpose is to explain that an August 11, 2015 letter[1] to Houston Independent School District from the organization Texas Appleseed regarding the district's discipline practices reflects a fundamental misunderstanding of the relationship between the frequency of suspensions from school and the proportion African Americans make up of persons suspended**.** In particular, consistent with understandings reflected in other materials created by Texas Appleseed, the letter reflects an understanding that reducing suspension rates among very young schoolchildren will tend to reduce the proportion African American boys make up of very young schoolchildren who are suspended. In fact, the opposite is the case. Reducing suspension rates among very young schoolchildren will tend to increase the proportion African American boys make up of very young schoolchildren who are suspended**.** A second purpose of this letter is to point out two interpretive problems in the Texas Appleseed letter relating to the relationship between the frequency of suspensions and student outcomes. These problems include the incorrect characterization of a November 2015 study as showing that a recent major reduction in discretionary suspensions in California correlates with higher district achievement. The study contained no such finding.

[1] To facilitate consideration of issues raised in letters such as this I include links to referenced materials in electronic copies of the letters. Such copies may be found by means of the Institutional Correspondence subpage to the Measuring Health Disparities page of jpscanlan.com.

A. Relationship Between the Frequency of Suspensions and the Proportion Disadvantaged Groups Make Up of Persons Suspended

With respect to its first purpose, this letter is closely related to three recent letters I have written regarding misunderstanding of the relationship between the frequency of an outcome and the proportions disadvantaged groups make up of persons experiencing the outcome. Two of those letters were sent to Texas Appleseed (Apr. 7, 2015) and McKinney, Texas Independent School District (Aug. 31, 2015). Those letters discuss analyses or reasoning of Texas Appleseed that reflects the mistaken view that generally reducing an adverse outcome will tend to reduce the proportion disadvantaged groups make up of persons experiencing the outcome. A third letter was sent to the Department of Health and Human Services and Department of Education (Aug. 24, 2015) (HHS/DOE Letter) and pertains to the "Joint Policy Statement on Expulsion and Suspension Policies in Early Childhood Settings" (Dec. 10, 2014) that the Texas Appleseed letter references in note 5 at page 3. The HHS/DOE letter explains to the agencies issuing the Joint Policy Statement that it contains the same mistaken understanding regarding the effects of reducing adverse school discipline outcomes on the proportion disadvantaged groups make up of persons experiencing the outcomes that has been reflected in analyses of Texas Appleseed.

To date, like the great majority of institutions and organization analyzing demographic differences in outcome rates,[2] Texas Appleseed's analyses have failed to recognize the 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, or the corresponding pattern whereby the rarer an outcome the greater tends to be the proportions groups most susceptible to

[2] The extent to which analyses of demographic differences are undermined by failures of understanding similar to that in the Texas Appleseed letter is reflected, among many other places, in my "Race and Mortality Revisited," *Society* (July/Aug. 2014) and my October 8, 2015 letter to the American Statistical Association. The bearing of such misunderstandings on the activities of particular institutions or organization to which I have written formal letters akin to the above-referenced letter to Texas Appleseed and the Departments of Health and Human Services and Education may be found in the following letters: Robert Wood Johnson Foundation (Apr. 8, 2009), National Quality Forum (Oct. 22, 2009), Institute of Medicine (June 1, 2010), The Commonwealth Fund (June 1, 2010), United States Department of Education (Apr. 18, 2012), United States Department of Justice (Apr. 23, 2012), Board of Governors of the Federal Reserve System (Mar. 4, 2013), Harvard University (Oct. 9, 2012), Harvard Medical School, Massachusetts General Hospital, et al. (Oct. 26, 2012), Senate Committee on Health, Education, Labor and Pensions (Apr. 1, 2013), Mailman School of Public Health of Columbia University (May 24, 2013), Investigations and Oversight Subcommittee of House Finance Committee (Dec. 4, 2013), Education Trust (April 30, 2014), Annie E. Casey Foundation (May 13, 2014), Institute of Medicine II (May 28, 2014), IDEA Data Center (Aug. 11, 2014), Education Law Center (Aug. 14, 2014), Financial Markets and Community Investment Program, Government Accountability Office (Sept. 9, 2014), Wisconsin Council on Families and Children's Race to Equity Project (Dec. 23, 2014), Portland, Oregon Board of Education (Feb. 25, 2015), Vermont Senate Committee on Education (Feb. 26, 2015), United States Department of Justice and City of Ferguson, Missouri (Mar. 9, 2015), Senate Committee on Health, Education, Labor and Pensions II (Mar. 20, 2015), City of Minneapolis, Minnesota (June 8, 2015), Agency for Healthcare Research and Quality (July 1, 2015), Chief Data Scientist of the Office of Science and Technology Policy (Sept. 8, 2015), House Judiciary Committee (Oct. 19, 2015), Boston Lawyers' Committee for Civil Rights and Economic Justice (Nov. 12, 2015).

the outcome make up of persons experiencing the outcome and the proportions such groups make up of persons avoiding the outcome.

The key points may be illustrated in Table 1 below, which is a replication of Table 1 (at 3) of the April 7, 2015 letter to Texas Appleseed. The table is based on a situation where the means of normal test score distributions of an advantaged group (AG) and a disadvantaged group (DG) differ by half a standard deviation and both distributions have the same standard deviation. For purposes of showing the effects of reducing failure rates on the proportion DG makes up of persons passing and persons failing the test, the illustration assumes that DG makes up 50% of test takers.

Given the difference between the underlying means, at the cutoff where 80% of AG passes the test, approximately 63 percent of DG would pass the test (with corresponding failure rates of 20% for AG and 37% for DG). The ratio of AG's pass rate to DG's pass rate (AG/DG Pass Ratio column) would thus be 1.27 while the ratio of DG's fail rate to AG's fail rate (DG/AG Fail Ratio column) would be 1.85.[3]

When the cutoff is lowered to the point where the pass rate for AG is 95%, the pass rate for DG would be approximately 87% (with corresponding failure rates of 5% for AG and 13% for DG). The ratio of AG's pass rate to DG's pass rate would thus decrease to 1.09 (from 1.27), while the ratio of DG's fail rate to AG's fail rate would increase to 2.60 (from 1.85).

The penultimate column shows that lowering the cutoff caused the proportion DG makes up of persons who pass to increase from 44% to 48%. The final column shows that lowering the cutoff caused the proportion DG makes up of persons who fail to increase from 65% to 72%.

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	DG Pass	AG Fail	DG Fail	AG/DG Pass Ratio	DG/AG Fail Ratio	DG Prop of Pass	DG Prop of Fail
High	80%	63%	20%	37%	1.27	1.85	44%	65%
Low	95%	87%	5%	13%	1.09	2.60	48%	72%

[3] While I commonly refer to patterns of relative differences in this letter, the table actually presents rate 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 the pattern by which relative differences tend to be affected by the frequency of an outcome.

The pattern whereby 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 is not peculiar to test score data or the numbers I used to illustrate it. Rather the pattern may be observed in the greater majority of situations where two groups differ in their susceptible to an outcome as discussed, among many other places, in my "Race and Mortality Revisited," *Society* (July/Aug. 2014) and my October 8, 2015 letter to the American Statistical Association. The pattern whereby reducing the frequency of an outcome tends to increase the proportions groups most susceptible to the outcome make up of persons experiencing the outcome and of persons avoiding the outcome is simply a corollary to the pattern of relative differences. Thus, whenever reductions in the frequency of an outcome are accompanied by increasing relative differences in rates of experiencing it, the proportions groups most susceptible to the outcome make up of persons experiencing the outcome will increase.[4]

So far as the published record reveals, recent general reductions in discipline rates have almost invariably been accompanied by increased relative racial/ethnic differences in discipline rates, as discussed on the following subpages of the Discipline Disparities page of jpscanlan.com (with jurisdiction indicated in title of the subpage): Los Angeles SWPBS, Denver Disparities, Florida Disparities, Maryland Disparities, California Disparities, Connecticut Disparities, Maryland Disparities, Minnesota Disparities, Rhode Island Disparities, St. Paul Disparities, Minneapolis Disparities, Beaverton (OR) Disparities, Portland (OR) Disparities, Montgomery County (MD) Disparities, and Henrico County (VA) Disparities. While the discussion of data regarding these jurisdictions is cast in terms of relative differences in adverse discipline outcomes, for reasons stated above, the increase in relative differences between the rates of two groups generally means an increase in the proportions groups most susceptible to the outcome make up of persons experiencing it. See also the letter to McKinney Independent School District (at 7) regarding the way that data referenced in a Texas Appleseed letter to that school district seems to indicate that a general reduction in student ticketing increased the proportion African Americans make up of students ticketed.

The same pattern may be found in the 2015 study referenced in note 3 (at 3) of the Texas Appleseed letter.[5] Though the study points out that African Americans experienced the largest declines in out-of-school suspension (OSS) rates in California between the 2011-12 and 2013-14 school years, such interpretation was based on percentage point changes. Endnote 2 (at 44) of the study indicates that relative differences between black and white rates increased.

---

[4] When there are only two groups, an increase in the relative difference between the rates at which the two groups experience the outcome necessarily means an increase in the proportion the more susceptible groups makes up of persons experiencing the outcome. The matter is a little more complicated when there are several groups. Observed patterns may also be affected by changes in the proportions various groups make up of the population potentially experiencing an outcome.

[5] Losen, D. et al., *Closing the School Discipline Gap in California: Signs of Progress*, Center for Civil Rights Remedies, Nov. 2015.

At page 4 of the Texas Appleseed letter, consistent with the way Texas Appleseed has appraised the size of demographic disparities in the circumstances discussed in the April 7, 2015 letter to Texas Appleseed and the August 31, 2015 letter to the McKinney, Texas Independent School District, the organization notes that African American boys made up 70% of very young schoolchildren who were suspended. Consistent with Texas Appleseed's mistaken understanding of the matter, the reference implies that reducing suspensions among very young schoolchildren should reduce this figure. But reductions in suspension rates short of eliminating all suspensions will tend to increase the figure.

Further, it should be recognized that the 70% figure cited in the Texas Appleseed letter – and like figures cited regarding the proportion African American students make up of suspended preschool students nationally, as in HHS/DOE Policy Statement – are functions of the fact that suspensions tend to be comparatively uncommon among young children. See Table 8 (at 342) of "Race and Mortality Revisited," which shows how, even though the strength of the forces causing black and white multiple suspension rates to differ is essentially the same in preschool as in K-12, the relative difference in multiple suspensions is larger in preschool than in K-12 because multiple suspensions are much less common in preschool than in K-12.[6]

B. Relationship Between Suspension Rates and Academic Achievement

The second purpose of this letter involves the Texas Appleseed letter's characterization of findings of the above-referenced November 2015 study pertaining to school discipline in California and the letter's reliance on a body of research concerning the relationship between suspensions and student outcomes.

The Texas Appleseed letter characterizes results of the November 2015 study of school discipline in California in the following manner (at 3, footnotes omitted):

> Research released in November, 2015, shows that a recent major reduction in discretionary suspensions in California correlates with higher district achievement. In the districts with reduced suspension rates, African American students experienced the most significant academic gains.

The findings of the study were not as characterized in the Texas Appleseed letter. The study examined patterns of changes in school suspension rates in all of California's school districts during a period of substantial overall declines in suspension rates. The study then specifically discussed two districts where African Americans, Latinos, and whites all experienced reductions in out of school suspension (OSS) rates between the 2011-12 and 2013-

[6] See the HHS/DOE letter (at 3-4) regarding the representations in the joint Policy Statement that preschool suspension rates are "high." That representation cannot be justified.

14 school years; where API scores improved between the 2011-12 and 2012-13 school years[7]; and where the African American and Latino API score gains were greater than the white gains.[8]

By my count based on data on enrollments and suspensions made available by the California Department of Education, there were 266 California school districts, where African Americans, Latinos, and whites all experienced declines in OSS rates between the 2011-12 and 2013-14 school years. Absent some general decline in API scores, one would expect to find within those 266 districts a substantial number of districts where API rates happened to improve for all groups and, within such districts, a substantial number where the African American and Latino scores happened to improve more than white scores.

The study did not attempt to determine whether there existed, or find, a correlation between reductions in OSS rates and higher district achievement. Nor did the study attempt to determine whether, of find, that in the districts with reduced OSS rates African American students experienced the most significant academic gains. Rather, the study simply reported that in two of the 266 districts where African Americans, Latinos, and whites all experienced OSS reductions African Americans and Latinos had larger API score increases than whites.[9]

The 2015 study did find a correlation between the suspension rates in particular districts with API scores in the districts. But such a finding is quite a different thing from finding a correlation between reductions in discipline rates in particular districts with changes in API scores in the districts. Cross-district studies of the correlations between suspension rates and API scores that, like the 2015 study, fail to adjust for student characteristics will invariably find that higher suspension rates are associated with lower API scores, just as studies of correlations between suspension rates and student poverty rates will invariably find that higher suspension rates are associated with higher poverty rates.

---

[7] The study (at 20-21) states that the two districts were chosen in part because the trends of reduced suspension rates and improved scores held for African Americans, Hispanics, and whites. But in one of the two districts the white rate remained constant.

[8] The statement in the Texas Appleseed letter that in the districts African Americans experienced the most significant academic gains is not correct. In one district (Berkeley Unified) the African American improvement was the largest. In other the other (Alameda Unified) the Latino gain was the largest. See the study's Table 12 at 22.

[9] The report itself states (at 20):

> Many school districts in California experienced a pattern of suspension rates falling while API scores rose. There are many reasons why this might occur, and sheer coincidence is always a possibility.

But it is not a question of whether it is a coincidence that many school districts had increasing API scores in conjunction with decreases in reductions in discipline rates. Rather, absent some strong pattern of general improvements or declines in API scores during this or any period examined, it would be remarkable not to find both many districts where API scores improved and many districts where API scores declined.

The 2015 study itself (at 1-2) recognized the cross-district correlations it examined were not probative of causality. The best could the study say about relationship between API scores and recent reductions in suspension in California was that there were no significant changes in API scores between 2013 and 2015 and that small decreases in average scaled-scores were similar to declines in most states including those without changes in discipline policies. In fact, the study expressed the concern that observed small decreases in API scores might be blamed on reductions in discipline rates.

Apart from the mischaracterization of the 2015 California study, the Texas Appleseed letter generally discusses a body of research regarding consequences of suspensions on student achievement. The Texas Appleseed letter can be considered to have fairly described the way such research is commonly characterized and arguably the actual conclusions themselves. But the part of the body of research with which I am familiar that addresses the consequences of stringent discipline policies, including the American Psychological Association (APA) Zero Tolerance study cited in notes 1 and 2 of the Texas Appleseed letter, is of extremely dubious validity. I discuss this issue with regard to the APA study on the APA Zero Tolerance Study subpage of the Discipline Disparities page. The criticisms of the APA study on that subpage would apply to essentially all statements in the Texas Appleseed letter regarding the correlations of suspensions with adverse student outcomes. For example, the study's conclusions fail to consider that students who engage in conduct that leads to suspensions tend to have characteristics associated with adverse educational outcomes irrespective of suspensions. Hence, such students will tend to have higher rates of adverse outcomes than other students regardless of the effects of suspensions or other disciplinary measures. That it may be difficult or impossible to determine whether disciplinary measures caused the adverse outcome rates to be higher than they otherwise would be is not a reason to regard the higher adverse outcome rates of disciplined students as meaningful evidence of the deleterious consequences of discipline practices.

Similar issues exist with respect to interpretations of data on correlations of frequency of suspensions with school wide academic achievement. Studies that purport to adjust for socioeconomic status and other factors invariably fail to do so adequately. Equally important, analyses of the relationship between suspension rates and student achievement fail to consider the extent to which high suspension rates are functions of higher rates of misconduct or the degree to which high rates of student misconduct contribute to and/or are functions of low student achievement.

Sincerely,

/s/ James P. Scanlan

James P. Scanlan

cc: Recipients and signatories of the Texas Appleseed Letter
Members of the Board of Directors of Texas Appleseed