

**Comment on Office for Civil Rights, U.S. Department of Education Request for  
Information Regarding the Nondiscriminatory Administration of School Discipline  
[Docket ID ED-2021-OCR-0068-0001]**

Submitted by:

James P. Scanlan  
1527 30<sup>th</sup> Street NW  
Washington, DC 20007  
jps@jpscanlan.com

## Introduction

A principal purpose of this comment is to explain that contrary to the belief promoted by the Department of Education and its Office of Civil Rights (DOE or OCR) and other federal agencies, generally reducing adverse school discipline outcomes tend to increase (a) relative racial differences in rates of experiencing the outcomes (a measure commonly presented in terms of the ratio of the Black rate to the white rate) and (b) the proportion Black students make up of suspended students. It is also to explain the consequences of DOE's leading the public and other arms of government to believe that policies will tend to reduce measures of racial disparity when in fact the policies will tend to increase those measures, especially in the context where the agency also promotes the belief, as it does in the instant Request for Information (Discipline RFI or RFI), that racial differences in adverse discipline outcomes are entirely result of racial bias. A related purpose of the comment is to show that the claim in the RFI that there are no racial/ethnic differences in disciplinable offenses is implausible and contradicted by many types of evidence.

Another related purpose of the comment is to explain how innumeracy at the DOE undermines its ability to evaluate the ways policies affect racial differences or racial bias with respect to any favorable or corresponding adverse educational outcome. As will be shown below, such innumeracy is reflected not only in the agency's promotion of a belief about the effects of reducing adverse discipline outcomes on certain measures of disparity that is the opposite of reality, but by a general failure to understand the ways all standard measures of differences involving outcome rates tend to be affected by the prevalence of an outcome. One notable example of that failure is the failure to understand that it is even possible for relative racial differences in favorable outcomes like test passage and relative differences in corresponding adverse outcomes like test failure to change in opposite directions as the prevalence the outcome changes, even though this will be the usual case. Another is the failure to understand even that lowering a test cutoff or generally improving test performance tends to increase relative differences between the failure rates of higher- and lower-scoring groups. Still another is the failure to understand that the relative difference an observer happens to be examining (i.e., in the favorable outcome or the corresponding adverse outcome) and the absolute difference between rates to change in the opposite directions as the prevalence of the outcomes changes, even though in the context of school discipline this will be the usual case. One failure of understanding that is especially pertinent to the Discipline RFI is the failure to understand that, while it may be possible to effectively quantify the differences in the circumstances of an advantaged and disadvantaged group reflected by their favorable or adverse outcome rates when those rates are actually known, it is impossible to effectively quantify such differences based on a comparison of the proportion the group makes up a of population and the proportion it makes up of persons experiencing an outcome, as is done in the RFI.

In the course of explaining these issues, this comment will also explain certain things about mistaken perceptions of the role of bias in causing outcome rates of advantaged and disadvantaged groups to differ,

This comment has three attachments. Attachment 1 is my December 8, 2017 [testimony](#), before U.S. Commission on Civil Rights (CCR) in which I explained that generally reducing adverse discipline outcomes tends to increase the aforementioned (a) and (b) for outcomes. The

testimony is discussed at various places in the U.S. Commission on Civil Rights' (CCR's) July 2019 report [\*Beyond Suspensions: Examining School Discipline Policies and Connections to the School-to-Prison Pipeline for Students of Color with Disabilities\*](#) (*Beyond Suspensions* report) on which the Discipline RFI placed significant reliance, and the testimony will be discussed further below. Attachment 2 is a [handout](#) I used to explain that matter and certain other matters regarding the measurement of demographic differences in educational outcomes at a March 22, 2018 meeting with DOE staff. Attachment 3 is my July 6, 2021 [response](#) to Office of Management and Budget Request for Information titled "Methods and Leading Practices for Advancing Equity and Support for Underserved Communities Through Government" (FR Doc No: 2021-09109). The response addresses a number of issues that are closely related to the issues addressed here, including the impossibility of quantifying a demographic difference based on the proportion a demographic group makes up of a population and the proportion it makes up of persons experiencing a favorable or adverse outcome.

In a great many places since 1987 I have explained that generally reducing an outcome tends to increase relative demographic differences in rates of experiencing the outcome while reducing relative differences in rates of avoiding the outcome. In explaining this issue, I have sometimes also explained how reducing an outcome tends to increase the proportion a group makes up both of persons who experience the outcome and persons who avoid the outcome. The failure to understand these patterns, however, is but part of a larger problem whereby persons analyzing demographic differences fail to understand how all standard measures of differences between outcome rates tend to be affected by the prevalence of an outcome, even when they do not have understandings of such effects that are the opposite of reality.

Discussions of varying comprehensiveness and complexity of both the larger problem and the issue that is the principal focus of this comment may be found in my "[Race and Mortality Revisited](#)," *Society* (July/Aug. 2014),<sup>1</sup> [Comments for the Commission on Evidence-Based Policymaking](#) (Nov. 14, 2016) (CEP Comments), [Letter to the American Statistical Association](#) (Oct. 8, 2015) (regarding measurement issue pertaining to a wide range of subjects); [Memorandum to HUD September 22, 2020 Expert Panel](#) (Sept. 19, 2020, updated Jan. 15, 2021) (HUD Panel Mem.) [Response](#) to Office of Management and Budget Request for Information titled "Methods and Leading Practices for Advancing Equity and Support for Underserved Communities Through Government" (OMB RFI Response) (Att. 3) (regarding measurement issues pertaining to a range of matters currently receiving great public attention); "[The Mismeasure of Health Disparities](#)," *Journal of Public Health Management and Practice* (July/Aug. 2016), [Letter to Harvard University](#) (Oct. 9, 2012), "[Measuring Health and Healthcare Disparities](#)," Federal Committee on Statistical Methodology 2013 Research Conference (regarding measurement issues pertaining mainly to health and healthcare disparities issues); 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. 17, 2014) (TDHCD Brief), and "[The Perverse Enforcement of Fair Lending Laws](#)," *Mortgage Banking*

---

<sup>1</sup> I have not attempted any consistency of formatting for the various references. In some cases, references are identified solely by links. A copy of this document that may be subsequently corrected or annotated is available [here](#). If it is corrected or annotated, such fact will be stated on the cover and corrections or annotations will be identified.

(May 2014) (regarding measurement issues mainly pertaining to fair lending issues); “[The Mismeasure of Discrimination](#),” Faculty Workshop, University of Kansas School of Law (Sept. 20, 2013) (Kansas Law Paper) (pertaining to identification and quantification of discrimination); and “[Measuring Discipline Disparities](#),” Testimony for U.S. Commission on Civil Rights Briefing “The School to Prison Pipeline: The Intersection of Students of Color with Disabilities” (Dec. 8, 2017) (Att. 1), and [handout](#) for meeting with DOE staff (Mar. 22, 2019 (Att. 2) (pertaining to quantifications of demographic differences in discipline and other educational outcome).

Many graphical and tabular illustrations of the pertinent patterns may be found in methods workshops given at [University of Massachusetts Medical School](#) (2015), [UC Irvine](#) (2015), [George Mason University](#) (2014), [University of Maryland](#) (2014), [University Minnesota](#) (2014), [Harvard University](#) (2012), and [American University](#) (2012). And many issues not addressed in any of the above works are addressed on the 100 plus pages and subpages of [jpscanlan.com](#) devoted to measurement/disparity issues, of which the most pertinent to the subject of Discipline RFI are the [Discipline Disparities](#) and [Educational Disparities](#) pages and their 66 subpages.

Most of this material is highly critical of all disparities research involving favorable and corresponding adverse outcome rates, commonly asserting or suggesting that such research has provided little of value but a great deal that is misleading about whether the forces causing adverse or favorable outcomes of advantaged and disadvantaged groups to differ are increasing or decreasing over time or are otherwise larger in one setting than another. The reader should recognize, however, that the more important it is to understand demographic differences – and how they are changing over time, whether they are larger in one setting than another, and the effects of policies on such difference – the more crucial it is to fully understand the issues addressed in these works. This is especially so when there is reason to believe that racial bias plays an important role in such differences, as explained in many of the works and especially in the discussion of Table 5 of “Race and Mortality Revisited.”

That reducing an outcome tends to increase relative differences in rates of experiencing an outcome, while reducing relative differences in rates of avoiding the outcome, is explained fairly simply, with a focus on mistaken beliefs that reducing adverse school discipline outcomes or adverse borrower outcomes by relaxing discipline and lending standard will tend to reduce relative racial differences in rates of experiencing the outcomes, in my “[Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies](#),” *Amstat News* (Dec. 2012), and [The Paradox of Lowering Standards](#),” *Baltimore Sun* (Aug. 5, 2013).” Both also discuss that the absolute (percentage point) difference between rates is the same (and thus directions of changes in such are the same) for an adverse outcome as for the corresponding favorable outcome, but that the absolute differences tends also to be affected by the prevalence of an outcome. The absolute difference, however, is affected by the prevalence of an outcome in a more complicated way than the two relative differences. For instant purposes, it suffices to say that for rates that are in the ranges where one typically finds Black and white out-of-school suspension rates (that is, well under 50% for both groups), general reductions in suspensions tend to reduce absolute differences between Black and white suspension rates, as was specifically noted in the *Baltimore Sun* commentary. Also, whereas relative racial differences in suspensions will commonly be

larger among female students than among male students and among students without disabilities than among students with disabilities (that is, where the outcome is less common), the absolute difference will almost invariably show an opposite pattern.

Both relative and absolute differences in adverse discipline outcomes will receive some attention here. Further, it is the failure to understand the way any measure of difference between outcome tends to be affected by the prevalence of an outcome that has prevented the DOE from saying anything useful about the way demographic difference in educational outcomes may be affected by policies, as I have discussed, for example, in [“Innumeracy at the Department of Education and the Congressional Committees Overseeing It.”](#) Federalist Society Blog (Aug. 24, 2017).

But it is the DOE’s promotion of the mistaken belief that generally reducing adverse discipline outcomes will tend to reduce the aforementioned (a) and (b) for the outcomes that has the most pernicious consequences. For one thing, promotion of that mistaken belief creates a situation where teachers and administrators who follow DOE guidance to relax discipline standards tend to increase the likelihood that they will be accused of discrimination or, in any event, accused of having especially large disparities. Presumably, some number of school principals or and school district superintendents have lost their positions because measures of racial disparity increased in the face of policies that were supposed to reduce those measures. Further, and perhaps most important, when measures of disparity increase in face of policies that the government has led the public to believe should reduce those measures, observers who believe that all or substantial parts of racial differences in discipline outcome are the result of discrimination will believe that racial bias must be increasing.<sup>2</sup>

The implications of certain of these consequences are heightened in circumstances where the government also leads the public to believe that all racial differences in adverse discipline outcomes are the result of racial bias, as the Discipline RFI attempts to do. That is, on page 8, the RFI adopts the finding from page 161 and the cover letter of the CCR’s *Beyond Suspensions* report that “[s]tudents of color as a whole, as well as by individual racial group, do not commit more disciplinable offenses than their white peers...” That language can only mean, or in any case will be taken to mean, that all racial differences in discipline outcomes are the result of racial bias.

In my view, the dissenting statement of CCR Commissioner Gail Heriot to the CCR’s *Beyond Suspensions* report (at 177-191) soundly refuted the report’s finding that there are no racial/ethnic differences disciplinable behavior. Data in some of the tables in Section A below also refute that claim.

---

<sup>2</sup> Observers have debated whether pressure to reduce measures of racial disparity in discipline will cause teachers and administrators to take race-conscious actions to reduce such measures (sometime cast in terms of causing quotas). It would seem obvious, however, that emphasis on statistical disparities, especially with a suggestion that bias plays a role in the disparities, will place pressures on decision-makers to take some form of race-conscious action even though they believe discipline decisions have been influenced by bias. That measures of racial disparity increase in the face of policies that are supposed them will almost certainly increase those pressures,

As suggested above, the statement that there are no racial/ethnic differences in disciplinable offenses has grave implications. And by making the statement the DOE is representing that it is true not only to the public but to other arms of government. Thus, the agency would wisely carefully consider whether it can responsibly leave the statement uncorrected. But, in the event that the agency does adhere to that claim that there are no racial/ethnic differences in disciplinable offenses, it should provide evidence to support the claim and provide guidance on whether and how belief in such claim will influence investigations of discrimination issues.

Further regarding differences in behavior, referencing the 2018 Government Accountability Office report titled “[Discipline Disparities for Black Students, Boys, and Students With Disabilities](#),” the RFI states: “This report found that Black students, boys, and students with disabilities were disproportionately disciplined ‘regardless of the type of disciplinary action, level of school poverty, or type of public school attended.’” The statement and other aspects of the RFI, at least with respect to students with disabilities, appear to regard differences in discipline rates between students with disabilities and students without disabilities and between male students and female students in the same way that it regards differences between discipline rates of Black students and white students.<sup>3</sup> But the RFI states nothing regarding behavioral differences between students with and without disabilities or between male and female students. Many would believe it obvious that at least students whose disabilities involve a behavioral disorder would have higher rates of conduct warranting removal from the classroom than other students (whether or not the disability is a basis for some sort of accommodation). And, based on their own schooling experiences, most people take for granted that male students engage in conduct warranting suspensions far more often than female students, something data on things like fighting would surely support.

Thus, the agency should clarify whether it in fact believes the observed differences between rates at which students with disabilities and students without disabilities are disciplined, or between rates at which male and female students are disciplined, should be appraised in a similar manner to the way differences between rates at which racial/ethnic groups are disciplined are appraised. It should also present whatever evidence might be available that provides reason to believe that observed differences in outcome rates between students with and without disabilities and male and female students are any greater than differences in behavior warrant.

\*\*\*

In the years since these the *Amstat News* and *Baltimore Sun* columns were published, all across the country general reductions in suspensions have typically been accompanied by increased relative racial differences in suspension rates, often with remarkable consistency. In the case of Maryland, whose Department of Education’s misunderstanding of the effects of relaxing discipline standards on relative racial differences in suspension rates was the focus of

---

<sup>3</sup> As reflected in the discussion in Section A.2.a, *infra*, RFI’s explanation of intersection of race and gender would appear to reflect a premise that absent some sort of bias, male and female suspension rates would be the same. That might be said of any of the comparisons between the proportion a racial/ethnic-gender subgroup makes up of students and the proportion it makes up of students experiencing some outcome, as has been common in recent OCR data summaries since 2016.

the *Baltimore Sun* commentary, a 2015 [study](#) by the Maryland Equity Project of the College of Education of the University of Maryland found that in 20 of the 23 Maryland school districts for which data were available, general reductions in public schools suspensions between the 2008-09 and 2013-14 school years were accompanied by increased relative differences between suspension rates of Black students other students. It also showed that statewide the ratio of the Black suspension rate to the white suspension rate had increased. A 2019 [study](#) funded by an entity funded by DOE and administered by Mathematica Policy Institute, although implying that general reductions in Maryland suspensions and expulsions between the 2009-10 and 2017-18 school years had made some progress in reducing the ratio of the Black rate of experiencing such outcomes to the white rate of experiencing the outcomes, provided data showing that the ratio of the Black rate to the white rate had actually increased from 2.6 (8.9%/3.7%) in the former school year to 2.9 (7.6%/2.6%) in the latter school year. The study also showed that the general reductions had been accompanied by an increase in the ratio of the rate for students with disabilities to the rate for students without disabilities from 2.1 (12.5%/6.0%) to 2.4 (8.9%/3.7%) over the period examined. See my “[Maryland Discipline Study Shows Usual – But Misunderstood – Effects of Policies on Measures of Racial Disparity](#),” *The Gunpowder Gazette* (Dec. 16, 2019). OCR State and National Estimations files show that in Maryland the ratio of the Black rate of one-or-more suspensions to the white rate of one-or-more suspensions increased from 2.39 (8.90%/3.73%) in the 2011-12 school year to 2.92 (7.24%/2.48%) in the 2017-18 school year. All of the above rates shown in parentheses would at the same time reveal that relative racial differences in rates of avoiding suspension (and absolute differences between rates) decreased during the periods examined, just as a numerate observers would expect.

An even greater example of consistency across school districts has been observed in in Minnesota. A Hechinger Report study described in “[In Minnesota, the discipline gap is not just an urban phenomenon](#),” MinnPost (Dec. 22, 2014) found that in all 73 Minnesota school districts that had data that could be analyzed, general reductions in suspensions between the 2010-11 and 2012-13 school years resulted in a larger proportionate decline for white students than Black students, which necessarily means relative difference between Black and white suspension rates increased.<sup>4</sup>

Many more examples where general reductions in suspensions were accompanied by increased relative differences in suspensions be found on the subpages of the [Discipline Disparities](#) page of [jpscanlan.com](#) bearing the names of states of local jurisdictions. I assume that in in the great majority of such cases, as in the 20 Maryland districts and 73 Minnesota districts, the absolute difference decreased.

---

<sup>4</sup> The report merely stated that the suspensions rates “dropped faster for white students than they did for black students in all 73 districts that had data available.” I assume that study found larger proportionate declines for white students than Black students, since it would be extraordinary to find that that even a substantial proportion of cases the white rate showed a larger absolute reduction than the Black rate. In any case, assuming white students had the lower original suspension rates, a larger absolute reduction for whites than Blacks would necessarily also mean a larger proportionate reduction for white students than Black Students. Probably in very high proportion of the 73 districts, however, the Black rate showed a larger absolute reduction than the white rate, with an accompanying decrease in the absolute difference between Black and white suspensions rates and a decrease in the relative difference between the rates of avoiding suspensions.

There have no doubt been many cases where general reductions in suspensions were accompanied by increased relative racial differences in suspensions rates (as in the 3 of 23 districts in Maryland), especially when general changes were small. But most reportage of situations where general reductions in suspensions were accompanied by decreases in racial disparities in suspensions have involved cases where those analyzing the data relied on absolute differences between rates to measure disparities. And the most prominent reportage regarding racial differences in suspensions that stated or implied that general reductions in suspensions had been accompanied by decreases in relative racial differences have involved situations where the relative difference actually increased. See [Massachusetts Disparities](#), [Virginia Disparities](#), [Allegheny County \(PA\) Disparities](#), [Denver Disparities](#), [Oakland \(CA\) Disparities](#) subpages of the Discipline Disparities page. The error apparently arose from the work of scholars who failed to understand that relative and absolute difference between Black and white suspension rates could change in opposite directions or that such is the usual case and was in fact the case in situations they analyzed.

As discussed in the Oakland and Denver pages, work of putative experts caused the school district themselves to mistakenly believe that restorative justice practices that generally reduced suspensions had reduced relative racial differences in suspensions. Those districts appear to have developed reputations for their expertise in addressing disparities issues precisely because of the failure of school officials to understand that the relative racial differences in their own districts had increased.

\*\*\*

These comments are informed to a degree by my viewing of the DOE May 11, 2021 webinar titled “[Brown 67 Years Later: Examining Disparities in School Discipline](#)” and attendance at a June 22, 2021 DOE virtual meeting regarding the Discipline RFI. The May 11 webinar gave substantial attention to the fact that relative racial differences in suspensions are greater among female students than male students, a subject that has received substantial attention lately in scholarly work and the media. According to my recollection, it was suggested in the webinar that the DOE would be giving substantial attention to such patterns and the intersectionality reflected in such patterns. I will give further attention to his subject in Section A with respect both to the concept of intersectionality that caused racial differences in suspensions to be perceived to be greater among female students than male students at the May 11 webinar and the concept of intersectionality that caused the Discipline RFI to highlight greater racial disparities in suspensions among male students than female students.

The most pertinent thing about the June 23 meeting was that attendees included University of Oregon Professor Kent McIntosh and Director of the UCLA Civil Rights Project Daniel Losen. Both have written extensively about discipline disparities issues and each presumably will respond to the RFI.

Professor McIntosh is a prominent member of the Positive Behavioral Interventions and Supports (PBIS) community, which, with DOE funding, has long promoted the view that generally reducing adverse discipline outcomes will tend to reduce the aforementioned (a) and

(b) for such outcomes (though various works from persons associated with that community have discussed discipline disparities and effects of policies on such disparities in terms of absolute differences between rates). Professor McIntosh was one of the four regional directors of the Technical Assistance (TA) Center on Positive Behavioral Interventions and Supports (PBIS) when, on October 1, 2018, the organization received a \$32 million dollar [grant](#) from the Center through 2023. He is also a member of the Center on PBIS Equity Workgroup and plays an important role in the sessions on equity issues at the annual [PBIS Leadership Forums](#).

Most significant, however, Professor McIntosh is the corresponding author a 2019 peer-reviewed paper on the measurement of discipline disparities co-authored by two other members of the Positive Behavioral Interventions and Supports working group. See Girvan et al., "[Tail, Tusk, and Trunk: What Different Metrics Reveal About Racial Disproportionality in School Discipline](#)," *Educational Psychologist* (2019). At least in its Figure 1, the paper appears to recognize that a general reduction in suspensions of the type that would result simply from lowering standards would tend to increase relative racial differences in suspension rates. The same figure recognizes that such reductions in suspensions would tend to reduce absolute differences in suspensions. That made the paper possibly the first the peer-reviewed paper by anyone in the education research community that has recognized that it even possible for the relative difference in an educational outcome and the absolute difference between rates for the outcome to change in opposite directions. The novelty of the understanding within the educational research community is underscored by the attention given to the work in Curran, F.C., [A Matter of Measurement: How Different Ways of Measuring Racial Gaps in School Suspensions Can Yield Drastically Different Conclusions About Racial Disparities in Discipline](#). *Educational Researcher* (May 2020).<sup>5</sup>

The paper also recognized the robustness of the measure of disparities in suspensions (or any other outcome) for comparing situations with different baseline rates that I employed and described in "Race and Mortality Revisited" and that I employed in both the CCR testimony and the March 2018 DOE handout and that I will use in certain tables below.

The paper would seem at least implicitly to recognize that simply causing general reductions in suspensions without affecting the forces causing Black and white rates to differ would tend to increase relative racial differences in suspension rates and the proportion Black students make up of suspended students. Since the publication of the paper, however, none of the PBIS presentation on equity issues with which Professor McIntosh has been involved (or any of the other equity focused presentation at such forums) appears to show any recognition of such pattern or to address whether any aspect of PBIS would be expected to counter such pattern. The online [material](#) from the forums suggest that any persons attending the conference who initially believed that reducing an outcome would tend to reduce relative racial differences in rates of experiencing the outcome and the proportion Black students make up of students experiencing the outcome would leave the forum with the same belief.

Mr. Losen has authored or co-authored a great deal of research regarding the effects of policies on racial disparities in adverse discipline outcomes, while analyzing those differences in terms of absolute differences between rates. I have discussed Mr. Losen's work in a number of

---

<sup>5</sup> This paper discusses some of the same Maryland data that I discussed in the Gunpowder Gazette post.

places, usually in the context of explaining that observers who have found general decrease in suspensions to be accompanied by decreases in disparities have, like Mr. Losen, quantified disparities in terms of absolute differences between suspension rates. See, e.g., [Relative Versus Absolute Differences](#) subpages of the Discipline Disparities page and the March 22, 2018 handout (at 3), as well as my [Letter to Antioch \(CA\) Unified School District](#) (Sept. 9, 2016). The last item (at 14) pointed out that Mr. Losen is one of the few persons who has recognized that is even possible to for relative and absolute differences in suspensions to change in opposite directions, though such recognition has failed to show an understanding of why that would be the usual case.<sup>6</sup>

Mr. Losen’s special significance in the instant situation lies in matters related to his participation at the U.S. Commission on Civil Rights December 8, 2017 briefing titled “The School to Prison Pipeline: The Intersection of Students of Color with Disabilities,” in which I presented my testimony explaining that reducing suspensions tended to increase, rather than reduce, relative racial differences in suspension rates. Orally at the briefing [Transcript](#) 52-53. I stressed the obligations of the Departments of Education and Justice to explain that their prior guidance regarding the effects of policies on measures of discipline was incorrect, as I had explained to the agencies and the Department of Health and Human Services by [letter](#) of July 17, 2017. But, noting that it was unclear whether the agencies would be able to understand the matter, I urged the Commission to take on that responsibility of understanding the issue as well as the larger problems in the analyses of the demographic difference discussed in the [CEP Comments](#) mentioned above.

At the conclusion of the panel in which I presented my testimony, Commission Chair Catherine Lhamon ([Transcript](#) 83) asked the other three panelists in the session, who apparently had been participants in generating federal discipline policy to provide a written response “about whether you believe that the generation of that policy would tend to reduce relative [(percentages)] [ ] racial differences in rates of experiencing discipline.” I do not know whether any of the panelists in fact responded.

But, on January 18, 2018, Mr. Losen, who was a witness on a later panel, submitted a [response](#) addressing the testimony of several witnesses including mine. In discuss in my testimony Mr. Losen began by stating as what he characterized as agreement with my technical observations (at 9): “Like James Scanlan, I have observed that the mathematical properties of suspension risk ratios do contribute to a tendency of such ratios to widen even if the absolute

---

<sup>6</sup> The letter to the Antioch Unified School District discusses Mr. Losen’s work throughout because of Mr. Losen’s involvement with the subject of the letter. The letter addressed a situation where a school district entered an agreement that was premised on the belief that generally reducing suspensions would tend to reduce the proportion Black students made up of suspended students. The agreement named Mr. Losen as the lead expert on discipline matters for determining compliance with the agreement. Thus, it involved a situation where, while general reductions in suspensions would tend to reduce the measures of disparity on which Mr. Losen commonly relied, it would tend to increase the measures of disparity on which the parties had been focused. The letter’s discussion of suspension data from the school district shown in the letter’s Table 2 (at 10), as well the letter’s discussion regarding characterizations of elements of disparities (at 10 n. 2), illustrate many of the issues both about measurement and about confusion in the descriptions of elements of a measure that are discussed in Section A infra.

racial differences narrow and suspension rates decline.” Thus, with respect Chair Lhamon’s crucial question, Mr. Losen agreed with my testimony.

Mr. Losen, however, went on to make certain statements that were cast in a manner as if disagreeing with my testimony. Most significant, Mr. Losen saw fit to explain why it was mathematically possible for the relative difference and the absolute difference to change in opposite directions.<sup>7</sup> The *Beyond Suspensions* report (at 144-45) then treated Mr. Losen’s response as if it in some in some manner countered my claim that general reducing suspensions tends to increase relative racial differences in suspension rates. But, the report, like Mr. Losen’s response, never actually said anything to call that claim into question.

Nevertheless, the *Beyond Suspensions* report issued in July 2019 went on to broadly promote the mistaken belief throughout that generally reducing suspensions, by means of PBIS and restorative justice programs and otherwise, would tend to reduce relative differences in rates of experiencing the outcome. One prominent instance of the report’s doing so may be found at page 148 where it cited as evidence of the continuing need for programs of the type recommended in the 2014 Dear Colleague letter that the ratio of the black suspension rate remained at 3.8 even after substantial reductions in suspensions between the 2011-12 and 2013-14 school years. It thus suggested that the reductions had made only limited progress in reducing the measure. In fact, however, as in other typical situations, the ratio was an increase from the 3.57 ratio in the earlier school year.

In any case, the DOE should not regard the *Beyond Suspensions* as actually calling into question the claim that generally reducing adverse discipline outcomes will tend to increase, rather than decrease, relative racial differences in rates of experiencing the outcome and the proportion Black students make up of students experiencing the outcome. And whether or not comments of either Professor McIntosh or Mr. Losen endeavor to provide any clarity on measurements issues, each is a useful resource for the agency to consult on whether reducing adverse school outcomes in fact tends to (and has commonly been observed to) reduce relative racial differences in rates of experiencing the outcomes.

Further, given the history of the government’s misleading the public on this matter, any discussion of effects of policies on racial differences in discipline rates that does not prominently disavow the government’s mistaken belief as to the effects of policies on measures of racial disparity further promote the mistaken belief. This was the case even with various actions of the last administration that were intended to reverse policies of the administration that preceded it but nevertheless reinforced the mistaken belief that reducing adverse outcomes would tend to reduce relative racial differences in rates of experiencing the outcomes. See “[COPAA v. DeVos](#)”

---

<sup>7</sup> Mr. Losen’s response showed little familiarity with my testimony or the work it references. He appeared to think that I maintained that the law requires the measurement of disparities in terms of relative difference in adverse outcome rates. I have for years maintained that the relative difference in either outcome is an unsound measure of association that should not be used for any reason. His response also pointed as if in opposition to my testimony that reducing suspensions tends to reduce relative differences in rates of not receiving suspensions. But that reducing an outcome tends reduce relative differences in rates of avoiding the outcome was specific point of my testimony and a principal point of almost everything I have written about the measurement of demographic differences.

[and the Government’s Continuing Numeracy Problem](#),” Federalist Society Blog (Sept. 12, 2019). And whether or not the RFI can be regarded as explicitly promoting that view, the failure to disavow it promotes the view just as effectively.

Section A discusses and illustrates patterns by which measures tend to be affected by the prevalence of an outcome with hypothetical and actual data, while also discussing pertinence of the data to perceptions about intersectionality, reasons to believe there are racial differences in disciplinable conduct and the drawing of inferences based on patterns observed in data.

Section B explains the impossibility of quantify demographic differences based on the proportion a demographic group makes up of persons a population and the proportion it makes up of persons experiencing an outcome.

#### **A. Explanation and Illustration of Patterns by Which Measure of Differences Involving Favorable and Corresponding Adverse Outcomes Tend to Be Affected by the Prevalence of An Outcome and Implications Respecting Issues Raised by the Discipline RFI.**

This section explains the problems with analyses of demographic differences that fail to recognize how the measure employed to quantify such differences tend to be affected by the prevalence of an outcome, with a principal focus on the mistaken belief that generally reducing outcomes like suspension from school will tend to relative racial differences in rates of experiencing the outcome (a measure commonly cast in terms of the ratio of the Black rate of experiencing the outcome to the white rate of experiencing the outcome). In illustrating the pertinent patterns with hypothetical and actual data, the section will also show particular problems with analyses of so-called intersectionality, either as the concept was employed in the May 11 webinar and in the CCR’s *Beyond Suspensions* report or as the concept is employed in the Discipline RFI. And it will use some of the data illustrating these patterns to show the implausibility of the claim that there are no racial/ethnic differences in disciplinable conduct and to show that one cannot draw inferences about processes based on the comparative size of a particular measure of differences between rates in different settings without understanding how the measure tends to be affected by the prevalence of an outcome in the different settings. While this section touches upon the mistaken belief that generally reducing an outcome for which two groups differ in their susceptibility of an outcome will tend to reduce the proportion the more susceptible group makes up of persons experiencing the outcome, a separate section will demonstrate the absurdity of attempting to quantify a demographic difference based on the proportion a group makes up of a population and the proportion it makes up of persons experiencing an outcome, as OCR has commonly attempted to do and as it does in the RFI.

Virtually all analyses of demographic differences involving favorable or corresponding adverse outcome have been undermined by a failure to understand patterns by which measures used to quantify demographic differences tend to be affected by the prevalence of an outcome. As a result of this failure, such analyses been unable to provide insight into whether the forces causing the outcome rates of advantaged and disadvantaged group to differ have increased or decreased over times or are larger in one setting than another.

The pattern by which measures tend to be affected by the prevalence of an outcome that is most pertinent to subjects of the RFI is that whereby the rarer an outcome the greater tends to be the relative difference between rates at which advantaged and disadvantaged groups experience the outcome and the smaller tends to be the relative difference between rates at which the groups avoid the outcome (i.e., experience the opposite outcome). The pattern could also be put in terms that whenever a favorable and corresponding adverse outcome change in prevalence, the relative difference for the increasing outcome tends to decrease while the relative difference for the decreasing outcome tends to increase. The pattern can easily be illustrated with test score data showing that lowering a test cutoff – and thus making test passage more common and test failure less common – tends to reduce relative difference between the pass rates of higher- and lower-scoring groups while increasing relative differences between the groups’ failure rates. The pattern can also be illustrated with myriad other types of data. It is also evident in what in fact commonly occurs when there is a change in the prevalence of a favorable and corresponding adverse outcome or the prevalence of an outcome varies from setting to setting, especially when that change or variation in prevalence is substantial. As discussed in [“Race and Mortality Revisited,”](#) *Society* (July/Aug. 2014), and [“The Mismeasure of Health Disparities,”](#) *Journal of Public Health Management and Practice* (July/Aug. 2016), the National Center for Health Statistics (NCHS) recognize this pattern more than 15 years ago in finding that improvement in health and healthcare tended to reduce relative racial/ethnic difference in the increasing favorable outcomes while increase differences in the corresponding decreasing adverse outcome.<sup>8</sup>

Nevertheless, even among persons who specialize in the analysis of demographic differences, virtually no one understands that it is even possible for the relative difference in a favorable outcome and the relative difference in the corresponding adverse outcome to change in opposite directions as the prevalence of an outcomes changes. And the overwhelming majority of such persons, and almost all federal agencies monitoring demographic differences for equity purposes, believe that generally reducing adverse criminal justice, school discipline, borrower, or health outcomes will tend to reduce, rather than increase, relative differences in rates of experiencing the outcomes. That belief is the exact opposite of reality.

While there are many reasons why reducing an adverse outcome may not increase relative demographic differences in rates of experiencing the outcome in a particular situation, no one has ever advanced a reason why one should expect that reducing the prevalence of an adverse outcome would usually reduce relative differences in rates of experiencing the outcome. Rather, researchers and policymakers have merely taken the matter for granted. And among the countless observers who reinforce the mistaken belief by pointing out that a relative difference in an adverse outcome increased or persisted “despite” a general decline in the outcome, none appears to have considered that the repeated departures from an expectation might be evidence that the expectation is unsound.

---

<sup>8</sup> As explained in those articles, however, NCHS was unable to act on this recognition in a sensible manner. While NCHS has never contradicted itself on the manner, it is not clear whether anyone within the agency today understands that the agency ever recognized this pattern. See my OMB RFI Response at 9-12 regarding the disarray in health and healthcare disparities research today.

## 1. Patterns in Normally-Distributed Test Scores

Table 1 below, a version of which also appears as Table 2 of my CCR testimony and in the dissenting statement of Commissioner Peter Kirsanow to the *Beyond Suspensions* report (at 215), shows the pass and fail rates of an advantaged group (AG) and a disadvantaged group (DG) at two cutoff points in a situation where the groups have normally distributed test scores with means that differ by half a standard deviation (a situation where approximately 31 percent of DG's scores are above the AG mean) and both distributions have the same standard deviation. The table also shows in the final columns measures of relative differences between the groups' rates of passing the test and relative differences between the groups' rates of failing the test.

**Table 1. Illustration of effect of lowering test cutoff on (a) relative difference between pass rates and (b) relative difference between failure rates of advantaged group (AG) and disadvantaged group (DG).**

Row	(1) AG Pass Rate	(2) DG Pass Rate	(3) AG Fail Rate	(4) DG Fail Rate	(5) AG/DG Pass Ratio	(6) DG/AG Fail Ratio
1	80%	63%	20%	37%	1.27	1.85
2	95%	87%	5%	13%	1.09	2.60

Column 5, which presents the ratio of AG's pass rate to DG's pass rate,<sup>9</sup> shows that at the higher cutoff, where pass rates are 80 percent for AG and 63 percent for DG, AG's pass rate is 1.27 times (27 percent greater than) DG's pass rate. If the cutoff is lowered to the point where AG's pass rate is 95 percent, DG's pass rate would be about 87 percent. At the lower cutoff, AG's pass rate is only 1.09 times (9 percent greater than) DG's pass rate. That lowering a cutoff tends to reduce relative differences in pass rates is well known in civil rights circles and underlies the widespread view that lowering a cutoff tends to reduce the disparate impact of tests on which some groups outperform others.

---

<sup>9</sup> While I usually refer to patterns of relative differences in this comment, the table actually presents ratios of two rates that I commonly refer to as rate ratio (and that are also termed risk ratios or relative risks). 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. In the former case, the larger the rate ratio, the larger the relative difference; in the latter case, the smaller the rate ratio, the larger the relative difference. It may be more common to employ the disadvantaged group's rate as the numerator for ratios regarding the favorable as well as the adverse outcome, which is the approach as to favorable outcomes of the "four-fifths" or "80 percent" rule for identifying disparate impact under the [Uniform Guideline for Employee Selection Procedures](#). I have sometimes employed this approach, as in "[Can We Actually Measure Health Disparities?](#)," *Chance* (Spring 2006). More recently, however, I have usually used the larger figure as the numerator for both ratios, in which case, as to both favorable and adverse outcomes, the larger the ratio, the larger the relative difference. Choice of numerator in the rate ratio, however, has no bearing on the patterns by which as the prevalence of an outcome changes, the two relative differences tend to change in opposite directions. I refer to patterns by which changes in the prevalence of an outcome affects relative differences rather than ratios because the accuracy of a statement as to how the prevalence of an outcome affects a ratio would turn on which figure is used as the numerator in the ratio.

But, whereas lowering a cutoff tends to reduce relative differences in pass rates, it tends to increase relative differences in failure rates. As shown in column 6, initially DG's failure rate was 1.85 times (85 percent greater than) AG's failure rate. With the lower cutoff, DG's failure rate is 2.6 times (160 percent greater than) AG's failure rate.

Thus, lowering the cutoff, and making test passage more common and test failure less common, decreased the relative difference in the increasing outcome but increased the relative difference in the decreasing outcome. The same results would be observed if, instead of lowering the cutoff, education were improved sufficiently to all enable all persons falling between the two cutoffs to reach the higher the cutoff. This pattern holds across the entire range of possible test scores.

One corollary to the pattern whereby the rarer an outcome the greater tends to be the relative difference in rates of experiencing the outcome and the smaller tends to be the relative difference in rates of avoiding the outcomes – and one of special pertinence to the universal misunderstanding of so called intersectionality in either the form discussed in the DOE May 11 webinar or the form that the RFI attempts to illustrate – may be described as follows: whenever a factor changes the prevalence of an outcome for two groups, or the prevalence of an outcome simply changes over time, the group with the lower baseline rates for the outcome tends to experience a larger proportionate change in its rate of experiencing the outcome while the other group tends to experience a larger proportionate change in its rate of experiencing the opposite outcome.<sup>10</sup> In the case of the lowering of the cutoff just discussed, AG would experience a larger proportionate decrease in its failure rate than DG (75% for AG versus 37.8% for DG), while DG would experience a larger proportionate increase in its pass rate than AG (22.2% for DG versus 18.75% for AG). As explained or illustrated in the longer paper and workshops listed in the introduction, anything that generally improves or worsens the situations of individuals regarding any matter, will commonly cause a larger proportionate change in adverse outcome rates for whites than Blacks while causing a larger proportionate change in the corresponding favorable outcome rates for Blacks than whites, just as it will do in the case of any advantaged group compared with any disadvantaged group where differences in baseline rates are substantial.

Here, too, however, few or no persons specializing in analyses of demographic differences are aware that such things are even possible. To my knowledge, no peer-reviewed paper in the social or medical sciences have ever shown such an awareness of such pattern even when rates discussed or presented in tables and figures of the paper make it obvious that the pattern exists in the data examined. The same may be said of persons analyzing data from clinical trials for purposes of determining how an intervention may affect different subgroups. See "Race and Mortality Revisited" (at 339-341).

---

<sup>10</sup> If this pattern seems in the least counterintuitive, it should be recognized that the group that has the higher baseline for the outcome being discussed necessarily has the lower baseline for the opposite outcome. Thus, the statement that the group with the lower baseline rate for an outcome tends to experience the larger proportionate change in the outcome necessarily implies that the other group tends to have larger proportionate change in the opposite outcome.

Table 2 (Table 3 of the CCR testimony) provides the same information as Table 1, while adding two columns to show the proportion DG makes up of persons who pass the test and persons who fail the test at the two cutoffs when DG makes up 50% of test takers.

**Table 2. Illustration of effect of lowering test cutoff on (a) relative difference between pass rates and (b) relative difference between failure rates of advantaged group (AG) and disadvantaged group (DG) and proportion DG makes up of (c) persons who pass the test and (d) persons who fail the test (where DG makes up 50% of test takers)**

Row	(1) AG Pass Rate	(2) DG Pass Rate	(3) AG Fail Rate	(4) DG Fail Rate	(5) AG/DG Pass Ratio	(6) DG/AG Fail Ratio	(7) DG Prop of Pass	(8) DG Prop of Fail
1	80%	63%	20%	37%	1.27	1.85	44%	65%
2	95%	87%	5%	13%	1.09	2.60	48%	72%

Columns (7) and (8) columns show that lowering the cutoff increases the proportion DG makes up both of persons who pass the test (from 44% to 48%) and persons who fail the test (from 65% to 72%). Because the proportion DG makes up of persons taking the test remains unchanged, lowering the cutoff would reduce all measures of difference between the proportion DG makes up of persons who take the test and persons who pass the test and increase all measures of difference between the proportion DG makes up of persons who take the test and persons who fail the test. That includes the absolute difference between the proportion a subject group makes up of the population and the proportion it makes up of persons experiencing the outcome that was used in the Government Accountability Office April 2018 [report](#) referenced in the RFI and the DOE’s April 2018 [OCR Summary](#) and the relative difference between said proportions that is employed to quantify demographic differences in the RFI.) While the hypothetical posits that DG is 50% of test takers in order to illustrate the pattern, the pattern of directions of changes in the proportions DG makes up of persons who pass the test and persons who fail the test would hold regardless of the proportion DG makes up of persons who take the test.

When there are only two groups in the population being examined, directions of change in measures of difference between the proportion DG makes up of the population and the proportion it makes up of persons experiencing an outcome as an outcome changes in prevalence are the same as the direction of the change in the relative difference for that outcome (as in the example). Even when there are only two groups in the population being examined, however, it is impossible to reasonably quantify a demographic difference based on a comparison between the proportion a group makes up a population and the proportion it makes up of persons experiencing an outcome, regardless of how one measures such differences. For, even then, such approach compares the situation of a subject group with the situation of the entire population, and the situation of that population is influenced by the rates of the two groups being compared and the proportion each group makes up of the total population, something that has nothing whatever to do with the comparative circumstances of the two groups that would be reflected by their outcome rates. Hence, such approach will typically make situations that are exactly the same as far the comparative circumstances of the two groups is concerned appear to be different with regard to such matter. It can also make a situation where the disparity between the rates of

two groups, however measured, is unquestionably smaller in one setting than another appear to be larger than in the other setting.

These are among the reasons why the presentation of demographic differences in terms of the proportion a group makes up of a population the proportion it makes up of persons experiencing an outcome is never useful and always in some manner misleading, and this holds regardless of whether or how one attempts to quantify the differences between said proportions. The same could be said of any comparison of a group's outcome rate with an overall outcome rate rather the rate of another group.

Because of the complexity of this subject, however, I defer discussion of the matter to Section B and will only touch upon the issue with respect to certain illustrations in this section. For purposes of the main points of this section, it suffices to say that the observations regarding, and illustrations of, the mistaken belief that generally reducing adverse discipline outcomes will tend to reduce relative differences between Black and white suspension rates apply as well to the mistaken belief that reducing the outcomes will tend to reduce the proportion Black students make up of students experiencing the outcomes.<sup>11</sup>

Table 3 is present the same situation shown in in Table 1, except a column has been added to show the absolute (percentage point) difference between the pass or fail rates of AG and DG at each cutoff. The column shows that reducing the cutoff reduced the absolute difference between those rates from 17 to 8 percentage points.

---

<sup>11</sup> I note at this point, however, that many people confuse the (a) proportion of a group that experiences an outcome (which is also the rate at which a group experiences the outcome) with (b) the proportion a group makes up of persons experiencing an outcome. As shown below, that is a reason the *Beyond Suspensions* would report that over 70% of Black students with disabilities received one-or-more out-of-school suspensions when in fact approximately 20% of such students received one-or-more suspensions. And countless observers who may be less familiar with data than the CCR staff may make similar mistakes by reading percentage in parentheses in the OCR tables that are described as "number and percentages of students receiving" some type of adverse or favorable outcome to mean (a) when in fact they mean (b). Some readers of the RFI may in fact read the instances where the RFI described a rate at which a group experienced an outcome to mean (a) when in fact it the statement are intended to mean (b).

The confusing of these proportions will be occurring even though any time there occurs a substantial reduction in the suspensions for two groups comprising a student population, the disadvantaged group's suspension rate will usually decrease while the proportion that comprises of suspended students will usually increase. And, while regardless of whether one is examining a favorable or adverse outcome, the rates at which the advantaged and disadvantaged groups experience the outcome tend usually to change in the same direction (especially when changes are substantial) the proportion the groups make up of persons experiencing the outcome must necessarily change in opposite directions. See the [Rhode Island Disparities](#) subpage of the Discipline Disparities page regarding a discipline disparities study reporting that the Black suspension rate had reached a high point over a particular period while the white suspension rate had reached a low point over that period. That would be an extraordinary occurrence if it had happened. But the data merely showed that the proportion each group made up of suspended students had reached highs and lows for the period. At least when there are only two groups in a student population (as was close to the case in the situation examined) and the proportions they make up of students does not change, if the proportion one group makes up of students experiencing an outcome reaches a high, the proportion the other group makes up of students experiencing the outcome must necessarily reach a low.

**Table 3. Illustration of effect of lowering test cutoff on (a) relative difference between pass Rates, (b) relative difference between failure rates, and (c) absolute difference between pass (or fail) rates of advantaged group (AG) and disadvantaged group (DG).**

Row	(1) AG Pass Rate	(2) DG Pass Rate	(3) AG Fail Rate	(4) DG Fail Rate	(5) AG/DG Pass Ratio	(6) DG/AG Fail Ratio	(7) Abs Df (PP)
1	80%	63%	20%	37%	1.27	1.85	17
2	95%	87%	5%	13%	1.09	2.60	8

But the information on the absolute difference in table should not be regarded as suggesting that lowering test cutoffs or reducing any adverse outcome will generally tend to reduce absolute differences between rates. The absolute difference between rates (and directions of changes in the absolute difference) are the same regardless of whether one examines the favorable or the adverse outcome. But the absolute differences is also affected by the prevalence of an outcome, though in a more complicated way than the two relative differences. Roughly, as uncommon outcome increase, absolute differences between rates of advantaged and disadvantaged group rates tend to increase, at least to the point where one group’s rate reaches 50%. As the outcome further increases, the absolute difference tends to decrease, at least after the point where both groups’ rates have reached 50%. Thus, the decrease in the absolute difference between the two rows in Table 3 is a function of the rate ranges shown in the table would not be found across a full range of test scores. Lowering a test score from a very high point to somewhat lower point would tend to increase absolute differences between rates.

Understanding the patterns by which absolute differences tend to be affected by the prevalence of an outcome is quite important to many things DOE does, since it and its contractors measure many demographic differences in educational outcomes in terms of absolute difference between rates. They invariably do so, however, without consideration of the extent to which observed patterns of changes in an absolute difference are functions of changes in the prevalence of an outcome. See my [“Innumeracy at the Department of Education and the Congressional Committees Overseeing It,”](#) Federalist Society Blog (Aug. 24, 2017). That means that such analyses have rarely said anything useful, though they have no doubt said much that is misleading, about the effects of a policy on differences in the circumstances of advantaged and disadvantaged groups respecting the outcome being discussed.<sup>12</sup> See generally each of the subpages on the [Educational Disparities](#) page of [jpscanlan.com](#).

I have discussed effects of changes in the prevalence of an outcome on the absolute difference between rates in all my longer works on the measurement of demographic differences since 2005, and, as discussed in the introduction, have often done so with particular focus on school discipline. But I did not discuss the absolute difference at all in the CCR testimony (except as implied by reference to the larger statistical issues I suggested that CCR address). The purpose of that testimony was to cause the CCR to understand, and to cause it to make other government agencies and the public understand, that reducing public school suspension tends to

<sup>12</sup> Even when a finding as to direction of change is broadly correct, as would be indicated, for example, when all measures change in the direction, the finding is misleading by suggesting the measure employed is effectively quantifying the demographic difference.

increase, not reduce, relative racial differences in suspension rates and the proportion Black students make up of suspended students. This is the matter regarding which agencies like the Departments of Justice and Education have neem been systematically misleading the public and other arms of government, with the pernicious consequences described in the introduction. To my knowledge, however, the agencies have not been misleading anyone with respect how reducing an adverse outcome tends to affect absolute differences between rates.

A principal purpose of this comment is to explain how changes in the prevalence of an outcome affect relative differences in rates of experiencing (or measures that are essentially functions of those relative differences). But, for a variety of reasons, it will be useful to include the absolute differences between rate in each of the illustrative tables that follow, and I will give the absolute differences between rates at least limited attention here.

As noted, the fact that lowering the cutoff reduces the absolute difference between rates in Table 1 is simply a function of the rates I chose to present in the table. But it warrant note that the adverse outcome (test failure) rates shown in the table are in the ranges usually found for many matters where demographic differences, including those involving school suspensions, arrests and incarceration, loan rejections, foreclosures and evictions, unemployment, and poverty, are commonly quantified in terms of relative differences in the adverse outcomes (rather than relative differences in the favorable outcome), and where the absolute difference tends to yield opposite conclusion with regard to the comparative size of disparity (whether over time or within one subgroup compared with another) from the relative difference in the adverse outcome. And increasingly observers are quantifying differences in certain of these matters in terms absolute differences between rates. These observers commonly reach opposite conclusions about the comparative size of a disparity from those relying on relative differences. Save in the small body of health and healthcare disparities research discussed in "Race and Mortality Revisited" and "The Mismeasure of Health Disparities," however, few researchers or commentators appear to recognize that it is even possible for this to happen, much less that it will almost always happen when the referenced adverse outcomes being discussed change substantially or are substantially larger in one setting than another (notable exceptions being Professor McIntosh and Mr. Losen, as discussed in the introduction).<sup>13</sup>

---

<sup>13</sup> The Federal Reserve is an entity that the public would generally expect to be especially expert in quantitative analysis. But even though income and credit score data make it obvious that relaxing lending requirement, while reducing relative racial differences in rates of meeting the standards, will tend to increase relative differences in rates of failure to meet the standards, for more than two decades the agency has operated under the belief that relaxing standards will tend to reduce relative differences in loan denial rates. Although the agency has often discussed various disparities issues in terms of absolute differences between rates, I have seen only one instance where employees of the agency have recognized that it even possible for relative and absolute differences to yield opposite conclusions. They did this with regard to unemployment rates in Cajner et al., "[Racial Gaps in Labor Market Outcomes in the Last Four Cycles](#)", Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C. (at 8 n.12). The manner in which the authors discuss the matter suggest that few economists in the agency understand the matter. And the authors' discussion of the way relative and absolute racial differences in unemployment changed in opposite directions during the Great Recession suggest that the authors have no idea that such pattern will almost certainly be found in every major recession and every major recovery.

To my knowledge, the Government Accountability Office, commonly regarded as the government's auditor, has never recognized that different measures may yield opposite conclusions about whether demographic differences are

When the actual rates are presented, that the relative difference for the adverse outcome and the absolute difference between rates yield opposite conclusions ought to be obvious about the comparative size of a demographic difference, at least to the numerate observer, and all observers ought to be able eventually to figure the matter out. But, as in the RFI itself, it is very common for persons and entities discussing demographic difference to report only the measure of difference on which they are relying – or only the proportions groups make up of the population and the population and the proportions the groups makes up of persons experiencing the outcome, with or without quantifying the difference between those two proportions – without stating the actual rates at which the groups being analyzed experience an outcome. Thus, observer will be unable to divine that the relative difference for the adverse outcome and the absolute difference between rates are yielding opposite conclusions about the comparative size of disparity (or that the relative difference for the favorable outcome yields an opposite conclusion from the relative difference in the adverse outcome and same conclusion as the absolute difference between rates).<sup>14</sup> See the OMB RFI Response (at 9-11) regarding the way the Agency for Healthcare Research and Quality’s failure to understand this matter has resulted in situations where the agency has highlighted in the National Healthcare Disparities Report as some of the fastest decreasing healthcare disparities over a particular period (on the basis of absolute differences between rates) situations where the agency would also regard the disparities to be much larger at the end of the period than at the beginning of the period (on the basis of relative differences in nonreceipt of care).

Before turning to illustrations of pertinent patterns with actual data, I note that there are several things one should keep in mind with regard to, and several things one may learn from, the simple test score illustration in Tables 1 and 3. First, DOE and other federal agencies that promote beliefs that reducing an adverse criminal justice, school discipline, or borrowing outcome will tend to reduce relative differences in rates of experiencing the outcome have not reasoned as follows: while it is true that lowering a test cutoff and thus decreasing test failure will tend to increase relative differences in test failure rates, there are reasons why one should not expect a similar result from reducing other adverse outcomes and, in fact, should expect reducing other adverse outcome to reduce relative differences in rates of experiencing those outcomes. Rather, despite decades of dealing with demographic differences in test outcomes, the agencies have yet to show an understanding even that lowering a test cutoff – or improving test performance or allowing the retaking of a test – will tend to increase relative differences between the failure rates of higher- and lower-scoring groups.

---

increasing or decreasing over time or about how policies may affect such differences. If it has yet recognized that, contrary to the belief the government promotes in many areas, reducing an adverse outcome tends to increase relative racial differences in rates of experiencing the outcomes – as I have explained in letters to the Comptroller General of [April 12, 2018](#), and [April 17, 2018](#) – the agency has not yet seen fit inform any arm of government.

<sup>14</sup> It is possible for all measures to change in the same direction, something that would be indicative of a meaningful change in the differences in the circumstances of the groups being compared. But anytime a relative difference for the outcome being discussed and the absolute difference yield opposite conclusions about the comparative size of a disparity, the unmentioned relative difference will necessarily yield an opposite conclusion from the mentioned relative difference and the same conclusion as the absolute difference.

Similarly, notwithstanding the connection between test scores and student proficiency, observers who analyze proficiency disparities in terms of relative difference in the favorable outcomes or in the corresponding adverse outcomes have yet to realize that general improvements or general worsening of education, or changing to an easier or harder test, will tend to cause relative differences in meeting standards to change in opposite directions from relative differences in failure to meet the standard. See the [CUNY ISLG Equality Indicators](#) subpage of the Education Disparities page of [jpscanlan.com](#).<sup>15</sup>

Thus, it would be useful to leadership of DOE and data analysts at DOE who do not yet understand that lowering a test cutoff tends to increase relative differences in test failure rates to attempt to understand such matter. Having come to understand such fact, those persons should consider the implications of such fact with respect to other matters where they have been operating under a belief that generally reducing an outcome will tend to reduce relative racial or other differences in rates of experiencing the outcome. But that would merely be a starting point for considering whether the agency can usefully analyze demographic differences regarding any educational outcome without considering the full implications of the patterns described in places like "Race and Mortality Revisited"

Second, neither the increase in the relative difference in rates of experiencing an adverse outcome caused by the decrease in prevalence of the outcome, nor the decrease in other measures caused by the decrease in the prevalence of the outcome should be read as indicating that disparity has increased or decreased in some meaningful sense. Rather, none of the measure is a sound indicator of the strength of the forces causing the outcome rates of the advantaged and disadvantaged group to differ because each tends to change solely because the prevalence of an outcome changes.

Further, it should be obvious that with respect to the favorable and adverse test outcomes shown in Table 1 through 3 – and any other situations where advantaged and disadvantaged groups' rates of experiencing a favorable or adverse outcome differ – the forces that cause the groups' favorable outcome rates to differ are exactly the same forces that cause the groups' adverse outcome rates to differ. Thus, for example, where the two rows reflect subjective judgments of decisionmakers and the question is whether there is greater likelihood of bias in one situation than the other, it makes no sense to say that one decisionmaker is more likely to be biased as to favorable outcomes and the other is more likely to be biased as to adverse outcomes. Nor would it make any sense to say that one decisionmaker is more likely to be biased in absolute terms while other is more likely to be biased in relative terms.

---

<sup>15</sup> The page discusses the way that [Equality Indicators](#) project of the [Institute for State and Local Governance](#) (ISLG) at the City University of New York measures proficiency disparities in terms of relative differences in nonproficiency rates and thus would reach opposite conclusions about the directions of changes in demographic differences regarding proficiency in New York City from those that New York City government's Center for Innovation through Data Intelligence would reach (based on relative differences in proficiency rates). It also discusses the way the ISLG provides guidance on measuring demographic differences in proficiency and other thing to other jurisdictions, most of which measure proficiency disparities in terms of relative differences in proficiency rates, thus tending to yield opposite conclusions about directions of changes from those ISLG would itself reach.

In fact, in the situation presented in the tables, or in situations where like rates exist with regard to some outcome in reality, there would be no rational basis for claiming the judgments causing the pattern in the first row are any more or less likely to be biased than the judgments causing the patterns in the second row. Similarly, there would be no rational basis for drawing an inference about the likelihood of bias based on the comparative size of any of the measures in the two rows, as, for example, where the adverse outcome rates in the first row are suspension rates where the appropriateness of suspension is more clear-cut and the adverse outcome rates in the second row are suspension rates in circumstances where the appropriateness of suspension is less clear-cut. And, in circumstances where the first row involves a subgroup where the adverse outcomes are more common and the second row involves a subgroup where the adverse are less common – as for example, where the adverse outcome in the first row are those of male students or students with disabilities and those in the second row are female students or students without disabilities – it should be obvious that there is value whatever in exploring the reasons for the comparative size of any of the measures from row to row without fully understanding the implications of the fact that the adverse outcome rates are generally higher in the first row than the second row. The same, of course, would apply to explorations of the reasons for the other standard measures to vary from row to row. Compare discussion of Table 2 in "Race and Mortality Revisited" (at 329-330, 343) involving the value of exploring reasons why measures of racial differences in poverty changed during the course of general change poverty in without consideration of what census data show would commonly happen any time there is a general increase or general decrease in poverty.

It is useful to keep in mind, however, that if the pass and fail rates are the results of a subjective judgments of different teachers, other things being equal, teachers who are more lenient graders or more effective teachers than other teachers will tend to show patterns more like that in Row 2 than Row 1, while other teachers will tend to show patterns more like that in Row 1 than Row 2. If the adverse outcome rates in the table are rates of suspension from school due to judgments of particular teachers, teachers who are more lenient disciplinarians, better able to maintain discipline without resort to suspensions, or more responsive to encouragements to generally reduce suspensions than other teachers will tend to show patterns more like that in Row 2 than in Row 1, while other teachers will tend to show patterns more like that in Row 1 than Row 2. Thus, other things being equal, decision-makers whose conduct most accords with that which governments encourage may face the greatest chances of being accused of discrimination on the basis of the comparative size of relative racial difference in adverse outcomes resulting from their actions. At least that will be the case so as long as the government and others persist in measuring racial/ethnic disparities in the way that they typically do.

Finally, any reduction in the forces causing the outcome rates of advantaged and disadvantaged groups to differ – including racial or other bias when that is a cause of such differences – will cause all measures of differences between outcome rates to be smaller than they would otherwise be. But it will be impossible to determine whether those force have increased or decreased over time, or the efficacy of policies aimed at reducing those forces, without understanding the ways measures of difference between outcomes rates are also being affected by changes in the prevalence of an outcome. For example, even though a program has reduced or eliminated the role of racial bias with respect to some outcome, general reductions in

the prevalence of an outcome may cause the relative racial difference to increase (though to a somewhat smaller degree than would have been the case without a reduction in bias).

## **2. Patterns in Data on Suspensions and Related Matters**

The rows of each of the following tables illustrate with actual data how the comparative size of various measures of differences between rates tend to be affected by the prevalence of an outcome in the same way that Table 3 demonstrates the point with hypothetical data. Each illustration will in some manner pertain to the mistaken belief of DOE and other federal agencies that reducing suspensions or other adverse discipline outcomes will tend to reduce, rather than increase, relative racial differences in rates of experiencing the outcomes. Certain of the tables will also show the mistaken significance attached to the comparative size of a measure with respect to different subgroups or different types of outcomes, whether with respect to so-called intersectionality or the drawing of inferences about processes. The data in certain of the tables will similarly show that the claim that there are no racial/ethnic differences in disciplinable conduct is manifestly implausible.

For simplicity, the tables only present the adverse outcome rates (leaving the corresponding favorable outcome rates to be inferred). And, in contrast to Tables 1 to 3, the disadvantaged group (Black) rate will be presented first and the column showing the relative difference in the adverse outcome will precede the column showing the relative differences in the adverse outcome. The final column shows the measure described in "Race and Mortality Revisited," and used in various tables of the CCR testimony, that is theoretically unaffected by the prevalence of an outcome. I commonly term the measure EES for estimated effect size and statisticians may refer to it as probit  $d$ . It is the measure that, as discussed in the introduction, the 2019 Educational Psychologist article by Professor McIntosh and colleagues recognized to be robust measure of discipline disparities. The approach derives from a pair of favorable or adverse outcomes the difference between means of hypothesized underlying normal risk distributions.<sup>16</sup> The figures in the column are the differences between those means in terms of percentage of a standard deviation. The favorable or adverse outcomes rates shown in tables 1 and 2 (or rather than unrounded figures that underlie the rounded figures shown in the table) would yield an EES value of .50 in both rows.

### **a. Illustration of patterns by which measures of differences between rates tend to be affected by the prevalence of an outcome with data on Black and white suspension rates among male students and female students with discussion of perceptions about the concept of intersectionality underlying discussions at the May 11 webinar and the concept of intersectionality reflected in the RFI.**

In a great many place since 1987, I have illustrated the pattern by which relative differences in adverse and corresponding favorable outcomes tend to be affected by the prevalence of an outcome by showing that in comparatively advantaged subpopulations, where adverse outcomes were comparatively uncommon, relative racial or other differences in the adverse

---

<sup>16</sup> The value is the same whether derived from the favorable outcome rates or corresponding adverse outcome rates, which would be a necessary feature of a sound measure of association (though not a sufficient feature as in the case of absolute difference between rates).

outcomes were larger, while relative racial or other differences in the corresponding favorable outcome were smaller, than in comparatively disadvantaged subpopulation where the adverse outcomes were more common. Table 1 of the [Discipline Disparities](#) page I created in 2012 used data on rates at which Black and white male and female public school students experienced certain levels of discipline to illustrate the point. Table 2 of that page similarly shows that among white students (where suspension is less common than among Black students), relative gender differences in suspensions are larger, while relative gender difference in avoiding suspensions are smaller, than among Black students.

Table 4 provides a similar illustration using out-of-school suspension rates of Black and white boys and girls that can be derived from OCR data for that 2017-18 school year.

**Table 4. Black and white rates of one-or-more out-of-school suspensions in by gender during 2017-18 school year, with measures of difference.**

Gender	Black Susp Rt	White Susp Rt	Bl/Wh Susp Ratio	Wh/Bl No Susp Ratio	Abs Diff (PP)	EES
Male	15.32%	4.85%	3.16	1.12	10.47	.637
Female	8.56%	1.64%	5.20	1.08	6.91	.766

Moving from the situation where suspensions are more common to the situation where they are less common, the table shows the usual pattern whereby relative differences in rates of experiencing the outcome increase while relative differences in rates of avoiding the outcome decrease. And, as would also typically be the case where outcome rates are in ranges one commonly finds for out-of-school suspensions of Black and white public school students, the absolute difference decreases.

While the table usefully illustrates the pattern by which certain measures tend to be affected by the prevalence of an outcomes, observers who attach any significance to the comparative size of any of the measures from row to row as it might bear on how race or factors associated with race may affect suspensions among female students differently from the way the factors affect suspensions among male students would be mistaken. The EES, however, does indicate that racial difference might be fairly regarded as somewhat greater among girl students than among boy students (which is consistent with that shown in the web page for the much earlier data that in the row where rates were principally functions of suspension rates). Whether the difference is such as to warrant funding of studies as why there is such a difference is another matter.

In recent years, however, observers have found something highly significant in the larger relative differences in suspensions among female students than male students. They do so while failing to understand that such is the expected pattern simply because suspensions are less common among female students than among male students and without showing any awareness that relative differences in avoiding suspension (and absolute differences between rates) are smaller among female students than male students. One example may be found in 2015 [study](#) by African American Policy Forum and the Center for Intersectionality and Social Policy Studies of Columbia Law School, which regarded the fact that the Black-white suspension ratio was larger

among girls than among boys as an example of intersectionality. This pattern was given much attention at the DOE May 11 [webinar 11](#), particularly in the presentation by Monique A. Morris, and, according to my recollection of the event, discussion at the webinar that the DOE should give increasing attention to intersectionality issues was focused on this issue.

I do not recall whether information was presented on actual suspension at the webinar. But one can take for granted that, regardless of the year, the actual rates would reveal the same pattern of larger relative differences in rates of avoiding suspensions (and larger absolute difference between rates) among male students than among female students shown in the table. And, to pursue further a point about presentation of data made earlier, the actual rates would enable astute observers to recognize that certain measures of differences between rates reveal patterns as to the comparative size of disparities that are the opposite of the patterns that are causing the perception of, and concern about, larger racial disparities among girls than among boys.<sup>17</sup>

As noted in the introduction Daniel Losen typically measures discipline disparities in terms of absolute differences between rates. In Losen, et al, [Are We Closing the Discipline Gap, Center for Civil Rights and Remedies](#) (Feb. 2015), Losen and colleague appraises the intersection of race and gender by comparing the percentage point differences between black and white suspension rates among male and female students, thus yielding opposite patterns from those discussed at the May 11 webinar and raising the question of why racial disparities are greater among male students than among female students.

Despite the attention given in the May 11 webinar to the intersectionality issue reflected in the seemingly larger racial disparities in suspensions among female than male students, the RFI introduces the concept of intersectionality by showing a seemingly larger racial disparity in suspensions among boys than girls. But this occurred for reasons quite different from those that caused Losen and colleagues to find larger racial disparities among boys than girls.

The RFI does this when it states (at 11):

Disparities worsen when you examine the intersection between race and sex.<sup>[18]</sup>  
According to the 2017-18 CRDC data, Black girls were the only group across all races or ethnicities for girls where a disparity in school suspensions was observed. Black girls

---

<sup>17</sup> A document termed “[Data Snapshot: 2017-18 National Data on School Discipline by Race and Gender](#),” published by the Georgetown Law Center on Poverty and Inequality and certain other entities, and which stresses the special vulnerability of girls of color compared with boys of color, presents the ratios of suspension and other adverse outcome by gender and race, but none of the actual rates. The highlighted summary statement on page 4 “This analysis of the 2017-18 data reveal that girls of color are at far higher risk of discipline in schools— more so than boys of color” is not correct. Many would understand it to mean merely that elevated risk, measured in relative terms, was greater among girls than among boys. But many would not. And few understand that the other measures of the comparative circumstances of Black and white students would yield opposite conclusions.

<sup>18</sup> The first sentence of this passage is a common catch phrase, like “disparities worsen when one digs deeper into the data.” The data that follow such statements, however, merely show that, however measured, some subgroup comparisons be larger than the overall difference while some will be smaller than the overall differences. The issue here, however, is somewhat more complicated.

accounted for 11.1 percent of in-school suspensions and 13.3 percent of out- of-school suspensions, which is almost two times their share of total student enrollment of 7.4 percent. Black boys accounted for 7.7 percent of total student enrollment and received both in-school suspensions and out-of-school suspensions at rates (20.1 percent and 24.9 percent, respectively) almost three times their share of total student enrollment -- the largest disparity across all race/ethnicity and sex groupings.

Here the RFI introduces a concept of intersectionality that differs from that employed in the May 11 webinar and in the Losen study, as well as in the *Beyond Suspensions* report, each of which are comparable to traditional subgroup analyses (also sometimes termed effect heterogeneity or interaction). This concept in the RFI seeks to quantify the cumulative effects of two aspects of identity in the manner can might be deemed comparable to comparisons of suspension rates of Black boys with disabilities with those of white boys without disabilities or, for a more extreme comparison, with the rates of white girls without disabilities. Such comparisons, while always yielding what are perceived to be very large disparities, do not shed light on any interactive effect of two factors.

In the case of male students, the RFI attempts to quantify the cumulative effect of two aspect of disadvantage with respect to the likelihood of suspension (Black race and male gender). But in the case of female students, the RFI attempts to quantify the effects one aspect of disadvantaged respecting likelihood of suspensions (Black race) and one aspect of advantage respecting likelihood of suspension (female gender). For reason suggested above and stated more explicitly in Section B, the presentation of disparities in terms of a comparison of the proportion a group or subgroup makes up the population and the proportion it makes up of persons experiencing an outcome both obscures what the RFI is actually doing and complicate the task of explaining it.

But the essence of what the RFI does in the quoted paragraph may be described as follows. Whereas a standard subgroup analysis would compare the Black female 8.56% suspension rate with white female 1.64% suspension rate, the RFI approach essentially compares the Black female 8.56% rate with a 4.45% overall rate (which is higher than 1.64% white female rate). One the other hand, whereas a standard subgroup analysis would compare the Black male 15.32% suspension rate with the white male 4.85% suspension rate, the RFI analysis essentially compares the Black male 15.32% rate with a 4.45% overall rate (which is lower than the 4.85% white boy rate).

The above discussion should not be read to mean that one ought not analyze data on racial differences by comparing rates at which various racial/ethnic group experience suspensions or any other outcome separately for male and female students. That, in fact, is preferred way to conduct such an analysis, and it would be essential but for the fact that among students of different racial/ethnic groups, boys and girls usually each make up approximately half of students. But, however one breaks down the data, it is impossible to plausibly quantify any difference – or draw any inferences about processes on the basis of said quantification – without understanding the way the measure employed tends to be affected by the prevalence of an outcome, much less while being completely unaware that is even possible for different

measures to show opposite patterns as to the comparative size of differences or the comparative size of effects.

**b. Illustration of patterns by which measures of difference between rates tend to be affected by the prevalence of an outcome with data on Black and white suspension rates among student with disabilities and without disabilities with discussion of the bearing of the data on perceptions about intersectionality and the mis-reportage of certain matters in the CCR’s Beyond Suspensions report.**

As suggested by the title of CCR’s December 2017 briefing (“The School to Prison Pipeline: The Intersection of Students of Color with Disabilities”) and the title of the CCR’s July 2019 report (“Beyond Suspensions: Examining School Discipline Policies and Connections to the School-to-Prison Pipeline for Students of Color with Disabilities”), and as indicated in the report itself at several places, the key purpose of the investigation leading to the report was to determine how racial differences in suspensions among students with disabilities may vary from racial differences in suspensions among students without disabilities. Ideally, CCR would have recognized at the outset that the relative racial differences in suspension rates would be larger, while the relative racial difference in rates of avoiding suspensions would be smaller, among students without disabilities than students with disabilities, simply because suspensions are less common among students without disabilities than students with disabilities. It could have verified that understanding by reference to any of the OCR State and National Estimations for the 2011-12 or 2013-14 school years. Such data would also have revealed the suspension rates for Black students with disabilities ranged from 17% to 21% during those years.

Instead, however, after an extensive investigation, the *Beyond Suspensions* report found that in the 2015-16 school year more than 70% of Black students with disabilities received one-or-more suspensions (rather than the true figure of 19.6%, as shown below). The report also indicated that relative differences in suspensions were larger among students with disabilities than students without disabilities, when in fact the opposite was the case.<sup>19</sup>

**Table 5. Black and white rates of one-or-more out-of-school suspensions among students with and without disabilities in the 2015-16 school year.**

Disability	Black Susp Rt	White Susp SR	Bl/Wh Susp Ratio	Wh/Bl No Susp Ratio	Abs Diff (PP)	EES
With	19.63%	7.10%	2.76	1.16	12.53	.613
Without	11.77%	2.62%	4.48	1.10	9.15	.753

---

<sup>19</sup> I did not specifically discuss this issue in my testimony, partly because my focus on explaining that reducing suspensions tend to increase certain measures of disparity and partly because I was uncertain as to what concept of intersectionality on the CCR had in mind. But commencing on page 6 of my testimony, I did illustrate the implications of reducing an outcome on measures of difference by means of comparisons of situations where suspensions were more common with situations where they were less common. Had I been more familiar with OCR data at the time, I might have thought to use information on Black and white suspension rates among students with and without disabilities as one of my illustrations (as in Table 5 here).

Table 5, which is based on the OCR State and National Estimation data for the 2015-16 school year (not available at the time of the release of the *Beyond Suspensions* report), shows the actual Black and white suspensions rates for students with disabilities and students without disabilities, along with the same measures of difference shown in the tables on suspension rates of Black and white male and female students.

The table shows the usual pattern whereby relative racial differences in suspensions are greater, while relative racial differences in rates of avoiding suspensions are smaller among the group where the suspensions are less common (i.e., students without disabilities compared with students with disabilities), a pattern that would almost certainly be found in every year for any jurisdiction with large numbers of students of both race with and without disabilities. The final column shows that, to the extent the matter can be effectively measured, the racial difference was greater among students without disabilities than students with disabilities.

Reordered, the rates in the table would also show that relative differences between the suspension rates of students with and without disabilities are greater, while relative differences in avoiding suspensions are smaller, among white students (where suspensions are less common) than among Black students (where suspensions are more common).

In contrast to recent prior years, the OCR State and National Estimations data for the 2018-19 school year does not appear to provide suspension data broken down by race and disability status). But it is virtually certain that, which respect to the comparative size of the each of the standard measures of differences between rates in two rows of data, the underlying data would show the same pattern as that shown for the 2015-16 school year in the table above. It is possible that variance in the EES values might be larger or smaller than that shown in the table, though it is improbable that patterns would be reversed.

CCR staff analyzed OCR raw data on suspensions for the 2015-16 school year. the *Beyond Suspensions* report described key data as follows (at 67) (footnotes indicated by omitted<sup>20</sup>):

While racial disparities in discipline rates are well documented,<sup>384</sup> data further show that these issues are compounded for students of color with disabilities. During the 2015 16 school year, 32 percent of black students with disabilities were suspended once, and almost 40 percent were suspended repeatedly, which mean these students were almost three times more likely to be suspended compared to white students with disabilities.<sup>385</sup> Further, multiracial boys with disabilities were also found to be seven times more likely to receive at least one out of school suspension compared to white students with disabilities.<sup>386</sup>

The 32% and 40% figures for rates at which Black students with disabilities received one suspension or multiple suspensions sum to a total of around 72% for one-or-more suspensions,

---

<sup>20</sup> Footnote 385 and 386 indicated that the relevant percentages calculated by Commission staff from OCR data collection for the 2015-16 school year.

rather than the 19.62% figure shown in the table. The error in the report apparently arises from the CCR staff's confusion of (a) the proportion Black students with disabilities who were suspended students made up of all students with disabilities who were suspended with (b) the proportion of Black students with disabilities who were suspended (i.e., the suspension rate of Black students with disabilities).<sup>21</sup>

The statement that Black students with disabilities were almost three times as likely to be suspended as white students (which appears to be correct and thus presumably based on the actual suspension rates) does not explicitly state that the relative racial difference was larger among students with disabilities than among students without disabilities. But, in context, such is the clear implication of the statement especially given that the main of the report was precisely to examine that issue. See report at 162. But as shown in the table above, and as one should expect, the relative racial difference in suspension rates was in fact larger among students without disabilities than students with disabilities.<sup>22</sup>

The pages following the quoted material say various things about the size of racial differences in suspension among students with disabilities as measured in terms of relative or absolute differences between rates.<sup>23</sup> But the report does not say anything about the size of racial differences in suspension among student without disabilities.

The RFI, as well as the *Beyond Suspensions* itself, may well invite research into the intersection of race and disability with respect to suspension and other types of outcomes, either according to the traditional concept of intersectionality underlying the discussion in the May 11 webinar or underlying the *Beyond Suspensions* report. But, as with any research into, or discussion of, intersection of race and gender, such research or discussion can provide nothing of value, though much that is misleading, if it fails to reflect an understanding of the implications of the prevalence of an outcome. Similarly, decisions about whether to fund any research should be informed by a full understanding, rather on the some seemingly provocative statistic that had caught the attention of a researcher typically, if not invariably, be unaware that is even possible

---

<sup>21</sup> The subsequently released State and National Estimations files for the 2015-16 school year show the proportion Black made up of all students receiving single suspensions and multiple suspension to be 34.7% and 41.8%, each of which is very close to the figures calculated by CCR staff (though apparently mistakenly identified as the suspension rates of Black students with disabilities rather than the proportions Black students made up of students with disabilities who were suspended).

<sup>22</sup> The statement that in the final sentence of the quoted paragraph that “multiracial boys with disabilities were also found to be seven times more likely to receive at least one out of school suspension compared to white students with disabilities” would seem also to be incorrect. OCR State and National Estimations data for the 2015-16 school year indicate that the rate of one-or-more suspensions for multiracial boys was 12.9%. Whether it was being compared with the rate for white boy students with disabilities (8.81%) or all white students with disabilities (7.10%), the multiracial rate would not have been more as much as two times as high as the white rate.

<sup>23</sup> As shown Table 5, and for reasons discussed, the absolute difference between black and white rates was larger among students with disabilities than students without disabilities. But apart from the discussion at pages 145-46 of the *Beyond Suspensions* report that provided the apparent reason for the report to disregard my testimony that generally reducing suspensions tend to increase relative differences in suspension rates, the report shows no awareness that relative and absolute differences between rates could provide opposite conclusions about the comparative size of disparities.

different measures to yield opposite conclusions. See page 3 of my July 17, 2017 [letter](#) to the Departments of Education, Health and Human Services, and Justice regarding the DOE’s review of the soundness of research it is funding.

In accord with the observations at end of Section 2.a, the above discussion should not be read to discourage breaking down discipline rates of racial/ethnic or racial/ethnic and gender groups by students with and without disabilities. In fact, it would be very useful (perhaps essential) to present comparisons of racial/ethnic group of racial/ethnic-gender group rates separately by students with and without disabilities, while breaking the data down further according to students with the types of disabilities where suspension rates are commonly high and the types of disabilities where suspension rates are more in accord with those of students without disabilities. That can shed light on factors that drive overall differences. Again, however, whether there is value in studying any sort of intersectionality believed to be reflected in such comparisons is another matter.

**c. Illustration of patterns by which measures of differences between rates tend to be affected by the prevalence of an outcome with data on Black and white suspension rates among students not living with and living with both birth parents, with discussion of the bearing of the data on perceptions about intersectionality and the implausibility of the claim in the RFI that there are no racial/ethnic differences in disciplinable behavior.**

The information in Tables 6 and 7 below is from a November 19, 2019 Institute for Family Studies article by Nichola Zill and W. Bradford Wilcox titled “[The Black-White Disparities in Suspensions: What is the Role of Family?](#)”

Table 6 is based on the article’s Figure 1, presented National Household Survey data on Black and white rates of suspension or expulsion for 2012 and 2016, years between which there were general declines in suspension or expulsion rates.

**Table 6. Black and white rates of suspension or expulsion from school for the years 2012 and 2016 with measures of difference.**

Year	Black Susp Rt	White Susp Rt	Bl/Wh Susp Ratio	Wh/Bl No Susp Ratio	Abs Diff (PP)	EES
2012	26.6%	9.7%	2.74	1.23	16.9	0.674
2016	23.4%	8.0%	2.93	1.20	15.4	0.679

The table shows the usual pattern of whereby general declines in suspensions were accompanied by an increase in the relative difference in suspension rates and a decrease in the relative difference in rates of avoiding suspensions (along with a decrease in the absolute difference between rates). The EES indicates that the forces causing Black and white rates to differ were essentially unchanged. The patterns are no different from that myriad situation of the type discussed in the introduction, but the information is worth presenting as a predicate to discussion of implications of information in the next table.

Table 7 is based on the Figure 3 or the Zill and Wilcox article, which show Black and white suspension rates in 2016 for students living with both birth parents and children living in other circumstances. This table, too, shows the usual pattern whereby in the subgroup where suspensions were less common, relative differences in the suspensions were larger, while relative differences in rates of avoiding suspensions and absolute differences between rates were smaller, than in the subgroup where the suspensions were more common. The EES indicates that forces causing Black and white rates to differ were essentially the same in the two subgroups.

**Table 7. Black and white suspension rates for students not living with both birth parents (BBP) and living with both birth parents.**

Student Living Situation	Black Susp Rt	White Susp Rt	Bl/Wh Susp Ratio	Wh/Bl No Susp Ratio	Abs Diff (PP)	EES
Not With BBP	28.0%	12.0%	2.33	1.22	16.0	0.590
With BBP	13.4%	4.9%	2.73	1.10	8.5	0.547

The data are of the usual type that might cause some to mistakenly think it worthwhile to explore why the relative racial differences is larger among students living with both birth parents than among students not living with both parents, or, as some might put it, why the benefits of living with both parents were not as great for Black students as white students.

Table 7, however, is also pertinent to the claim in the RFI and the *Beyond Suspensions* report that there are no racial/ethnic differences in disciplinable behavior, which is how the authors regarded the information in the figure. One of the most obvious reasons why such a claim is implausible is that disadvantaged racial/ethnics groups are commonly disproportionately represented in the categories defined by parental circumstances, socioeconomic status, and academic achievement where disciplinable behavior rates are higher for all racial/ethnic groups than in other categories. Figure 2 of the Zill and Wilcox article importantly shows that 72% of Black students and 46% of Hispanic students, compared with 37% of white students and 26% of Asian students, do not live with both birth parents. While the suspension rates are only indirect indicators of disciplinable behavior, absent reason to believe that schools are systematically discriminating against students who do not live with both birth parents, the information in the provide sound evidence that there are racial difference in disciplinable behavior.

Where data are available, similar patterns of higher suspension for all racial/ethnic groups within disadvantaged populations defined by socioeconomic status or academic achievement would be also be shown. Indeed, most people who attended 12 or so years of elementary and secondary educations would attest to such pattern, whether phrased in terms of difference in suspensions or difference in behavior of a type that might lead to suspension if observed by a teacher. And that would be so whether or not the person observing the pattern had any knowledge of racial/ethnic differences in falling into categories and whether the persons were educated in schools where all students were of one racial/ethnic group or schools where students were of many racial/ethnic groups.

Observers who seek to minimize the role of differences in behavior in suspension rate differences would point to the fact that racial differences exist within both subgroups or that

adjustment for racial differences do not fully eliminate the racial differences.<sup>24</sup> Both statements are true,<sup>25</sup> though such statements are often phrased in terms that lead others to believe there are no racial/ethnic differences in disciplinable behavior.<sup>26</sup> But the fact remains that data like that in Table 7 are among the many types of data showing that the statement there are no racial/ethnic differences in disciplinable behavior is implausible.

The best way to adjust for differences in family living circumstances given the information in Table 7 is to estimate what the Black suspension rate would be if Black students had the same family structure as white students. That approach leaves the white overall suspension rate the same as the 8.0% figure for 2016 shown in Table 6, while the overall Black suspension rate is reduced from 23.4% to 18.0%.<sup>27</sup> Thus, all measures of disparity would be reduced from those shown for 2016 in Table 6.<sup>28</sup>

Observers may debate the extent to which factors other than those reflected by family structure (and other than discrimination) might account for racial differences in suspensions. But the debate should be about differences that remain after adjustment for available data not about differences without adjustment (and with recognitions of the inadequacies of data used for the

---

<sup>24</sup> In the lending disparities context, observers sometimes attach special significance to the fact that the relative racial difference in adverse outcomes are even larger within among advantaged subpopulations than disadvantaged subpopulations. See my [The Perverse Enforcement of Fair Lending Laws](#) (Mortgage Banking, May 2014)

<sup>25</sup> See the [Underadjustment Issues](#) and [Disparate Treatment](#) pages of [jpscanlan.com](#) regarding the way that even good faith efforts to adjust for background characteristics rarely if ever fully adjust for difference in such characteristics. The efforts that fail to acknowledge the inadequacies of available data or the way such inadequacies tend to systematically undermine the explanatory value of a characteristic are always open to question as to expertise or objectivity.

<sup>26</sup> The [January 2014 Dear Colleague Letter](#), after describing relative racial differences in suspensions of differences or between the proportion Black students make up of students and the proportion they make up of disciplined students, states: “However, research suggests that the substantial racial disparities of the kind reflected in the CRDC data are not explained by more frequent or more serious misbehavior by students of color.” (footnote omitted). This statement is literally true in the sense that research typically fails to show that difference in behavior or behavior-related characteristics fully explain observed differences. But it is also a statement that might lead some to believe there are no racial differences in behavior. In any case, it is materially different statement from the statement in the *Beyond Suspensions* report and the RFI that there are no racial/ethnic differences in disciplinable behavior.

<sup>27</sup> The Zill and Wilcox article employed logistic regression to adjust for characteristics while including rates for all racial/ethnic groups within the group analyzed. That approach allows, which yields estimate of what the Black and white rates would be if they had the same characteristics as the overall population, allows the adjustment to be influenced by the both the rates of all groups other than Blacks and whites and the proportion every group makes up of the total. Even if the group analyzed were comprised solely of Blacks and whites, the proportion Black students make up of total students would affect the adjustment in ways that have nothing to do with the crucial question. The problem may be compared to the problems illustrated in the tables of Section B.

<sup>28</sup> The 18.0% and 8.0% rates yield an EES figure or .489 (compared with the .69 figure shown in the second row of Table 6. The .489 figure reflects a difference that is approximately the same as the reflected in both rows of Tables 1 to 3.

adjustment). It should not be informed by an implausible belief that there are no racial/ethnic differences in disciplinable behavior.<sup>29</sup>

**d. Illustration of patterns by which measures of differences between rates tend to be affected by the prevalence of an outcome with data on Black and white self-reported rates of fighting or indicators of more violent fighting, with discussion of the bearing of the data on the implausibility of the claim in the RFI that there are no racial/ethnic differences in disciplinable behavior.**

Table 8 below is a variation on Table 2 of my [“The Misunderstood Relationship Between Racial Differences in Conduct and Racial Differences in School Discipline and Criminal Justice Outcomes,”](#) Federalist Society Blog (Dec. 20, 2017). The Black and white adverse outcome rates in the two rows are self-reported rates of engaging in fights or being injured in fights according to a CDC survey reported in 2016. The data were used in the article to explain why a certain ratio of the Black rate of fighting to the white rate of fighting explain a higher ratio of rates for being punished or arrested for fighting, given that more severe discipline tend to be limited to the less common, and more violent, cases of fighting (rates of which would be more in line with rates of fights that resulted in injury).

**Table 8. Rates at which Black and white students reported fighting and being injured fights, with measures of difference.**

Event	Black Adverse Rate	White Adverse Ratio	Bl/Wh Adverse Ratio	Wh/Bl Favorable Ratio	Abs Diff (PP)	EES
Fighting	32.4%	20.1%	1.61	1.18	1.61	0.38
Injured in Fight	4.7%	1.9%	2.47	1.03	2.47	0.40

The table shows the usual patterns as to the comparative size of the two relative differences and the absolute difference for the more common outcome (fighting) and the less common outcome (fighting with injury). The EES indicates that the forces causing the rates to differ are essentially the same. In fact, based on the Black and white rates for engaging in fights, and either the Black (or white) rate for engaging in fights that led to injury, one can closely predict the corresponding white (or Black) rate of engaging in fights that resulted in injury.<sup>30</sup>

<sup>29</sup> See my [“The Misunderstood Relationship Between Racial Differences in Conduct and Racial Differences in School Discipline and Criminal Justice Outcomes,”](#) Federalist Society Blog (Dec. 20, 2017), regarding what I believe is the best evidence of the maximum contribution of racial bias to observed racial differences in discipline rates.

<sup>30</sup> It may be that, on average, persons injured in fights are less often the aggressors in fights and hence less likely to be punished for the fight in which in the injury occurred. That does not diminish the probative value of the information, given that participants in fights will usually be of the same race. Thus, rates at which students of a racial/ethnic group are injured in fights are comparable to rates of at which students of a racial/ethnic group injure others in fights.

A similar pattern is found in responses of persons between ages 12 and 17 to National Survey on Drug Use and Health (NSDUH) reported by Maynard, et al.<sup>31</sup> Table 9 shows the Black and white rates of reporting fighting at school or at work or attacking someone with an intent to injure.

**Table 9. Rates at which Black and white youths ages 12 to 17 reported fighting and attacking person with intent to injure, with measures of difference.**

Event	Black Adverse Rate	White Adverse Ratio	Bl/Wh Adverse Ratio	Wh/Bl Favorable Ratio	Abs Diff (PP)	EES
Fighting	29.0%	18.7%	1.55	1.15	10.3	0.34
Attack with Intent to Harm	11.9%	5.8%	2.05	1.07	6.1	0.39

The table shows the same pattern as the preceding tables with respect to the standard measures of differences between rates for the less serious and more common types of physical altercation and the more serious and less common type of physical altercation. The difference between the EES values in the two rows remains small, though arguably slightly larger than in the prior table.

Importantly, however, both rows of both tables also directly show that rates at which Black and white students engage in types of behavior that school administrators would reasonably regard as basis for suspensions or other disciplinary actions are not the same.

The rates in the tables may be compared with rates at which in 2015 Black and white students in grades 9-12 said they were in a fight on school property in the last 12 months, respectively 12.6% and 5.6%, discussed in Commissioner Heriot’s dissenting statement in the *Beyond Suspensions* report at 180. While each of the other measures of differences between rates falls between the values in the values in the two rows of Tables 8 and 9, the EES derived from the 12.6% and 5.6% would be .444, slightly higher than the figures shown in each row of the tables. In any case, that information is further evidence that Black and white rates of engaging in disciplinable behavior are not the same. There is substantial further such information in Commissioner Heriot’s dissenting statement.

**e. Illustration of patterns by which measures of differences between rates tend to be affected by the prevalence of an outcome with data on Black and white self-reported rates of skipping school or cutting class, with discussion of the bearing of the data on the implausibility of the claim in the RFI that there are no racial/ethnic differences in disciplinable behavior.**

Table 10 below is based on data from a 2016 article on truancy and suspension in Los Angeles (Nadra Nittle, [“Truancy, suspension rates drop in greater Los Angeles area schools,”](#)

<sup>31</sup> Maynard et al., [Truancy in the United States: Examining Temporal Trends and Correlates by Race, Age, and Gender](#). Child Youth Serv Rev. (Aug. 8, 2017).

LA School Report (Mar. 7, 2016)). The article presented California Department of Education data on rates at which public school students of different racial/ethnic groups in grades 7, 9, and 11 reported five levels of frequencies. The table shows rates of reaching each of those levels for Black and white students, with the same measures of difference presented in previous tables.

**Table 10. Rate at which Black and white California public school students in grades 7, 9, and 11 reported various levels of skipping school or cutting class in the past 12 months, with measures of difference.**

Frequency	Black Rate	White Rate	Bl/Wh Adv Ratio	Wh/Bl Fav Ratio	Abs Df (PP)	EES
Any	34.5%	31.9%	1.08	1.04	2.6	0.07
More than twice	21.7%	16.2%	1.34	1.07	5.5	0.25
More than a few times	9.6%	6.0%	1.60	1.04	3.6	0.33
More than monthly	7.2%	3.7%	1.95	1.04	3.5	0.21
More than weekly	5.6%	2.1%	2.67	1.04	3.5	0.45

The table shows the usual pattern whereby as outcome becomes less common, relative racial differences in rates of experiencing the outcome increase. It does not, however, show the usual pattern with regard to the other indicators, though such pattern may be found to some degree in the second through fifth rows, which contain data on what might be deemed greater than rare truancy. The departures may have something to do with the fact that the data are from many jurisdictions and patterns that may found within each jurisdiction may not always be found in aggregated data from many jurisdictions. The departures may also reflect the fact that single instances of some form of truancy may involve factors with respect to which there may be little variation across racial groups. But whatever the reasons for this departure, the data nevertheless show the following things that are pertinent to issues addressed in this comment.

First, the data show that increasing the level of truancy required to trigger discipline for truancy will tend to increase relative racial differences in punishment for truancy. Second, it shows that, with regard to this type of behavior that schools may regard as warranting some discipline, especially in more extreme cases, Black and white rates are not the same. Third, inasmuch as frequent truancy is a factor that would typically be associated with other types of disciplinable behavior, it provides further evidence of racial differences respecting those other types of disciplinable behavior.

Further with respect to truancy, it warrants note that the January 2014 Dear Colleague itself suggest that disadvantaged minorities may be more often truant than other students, while suggesting that, because of such differences, punishments for truancy must be justified. But the in statement the *Beyond Suspensions* report and Discipline RFI respecting the absence of racial/ethnic difference in disciplinable behavior, with its implication that racial bias is the entire reason for observed racial differences in suspensions, says nothing about behavior for which discipline can be justified. So, the Dear Colleague letter itself strongly implies that there are racial differences in truancy, an implication borne out by the data above and presumably by all sorts of other data that are readily available to the DOE.

**f. Illustration of patterns by which measures of differences between outcome rates tend to be affected by the prevalence of an outcome with data on probabilities that Black and white students engaging in disruptive conduct or fighting without injury are suspended, with discussion of the patterns regarding drawing inferences about processes based on the comparative size of relative racial differences in suspensions for different types of behavior.**

Table 11 is based on data from March 2018 Community Center for Education Results report titled “Discipline Practices in South Seattle and King County Schools Center.” The Black and white rates are from the table on page 20 of the report showing probability of suspensions for white students found to have engaged in fighting without injury (where rates of suspension for students found to have engaged in such conduct were quite high) and for disruptive conduct (where suspensions rates for students found to have engaged in such conduct were much lower).<sup>32</sup>

The table show the usual pattern whereby in the situation where the outcome is less common, the relative racial difference in experiencing the outcome is larger while the relative difference in rates of avoiding the outcome is smaller, than in the situation where the outcome is more common. The table also show the pattern shown in the other tables whereby the absolute difference is larger in the second row than the first. One should recognize, however, that, in contrast to the earlier tables, in this case the rates in both rows are not in ranges where reductions in the adverse outcome will tend to reduce absolute differences between rates of experiencing the outcome. The rates for fighting without injury are in ranges where reductions in suspension rates would tend to increase absolute differences between rates. That would also be the case with respect to what are presumably the still higher rates of suspensions for fights that cause injuries.

The EES indicates that the forces causing Black and white suspension rates to differ are actually substantially smaller for disruptive conduct than for fighting.

**Table 11. Probabilities at which Black and white students who engaged in fighting without injury and disruptive conduct will be suspended in schools in Seattle areas, with measures of difference.**

<b>Offense</b>	<b>Black Rate</b>	<b>White Rate</b>	<b>Bl/Wh Susp Ratio</b>	<b>Wh/Bl No Susp Ratio</b>	<b>Abs Df (PP)</b>	<b>EES</b>
Fighting (w/o injury)	76%	55%	1.38	1.88	21.00	.580
Disruptive Conduct	18%	10%	1.80	1.10	8.00	.366

---

<sup>32</sup> Rates at which students found to have engaged in certain conduct are suspended are different from rates at students will be found to have been suspended for the conduct. Rates of disruptive conduct are presumably much higher than rates of fighting. Thus, even though the rates at which students engaging in disruptive conduct are suspended are lower than rates at which students engaged in fighting are suspended, a higher proportion of students will be found have been suspended for disruptive conduct than for fighting.

The fact that relative racial differences for subjectively-identified (commonly less serious) offenses than objectively-identified (commonly more serious) has played importantly into arguments that racial bias is an important cause of racial differences in discipline decisions. The same argument could be based on the comparative size of the relative differences in suspensions shown in the two rows of Table 11. But, as is the common or usual case, the comparative relative difference in decision not to impose discipline would support a contrary interpretation. And, as noted, the EES indicates that racial difference is actually much smaller for the less serious offense.

The aforementioned argument regarding subjectivity commonly goes to determinations of whether a student has engaged in a particular type of conduct. That is, all that is known are suspension rates at which students who are suspended are suspended for particular types of conduct, not rates at which students who have been identified as engaging in particular type of conduct are suspended (as in the data in the table above). But, as explained in my [Offense Type Issues](#) webpage, the same statistical issues are pertinent. That is, the more a sanction is limited to the most extreme instance of a certain type of conduct – as will typically be the case for types of conduct falling into a category generally considered less serious compared with a type of conduct falling into a category generally considered more serious – the greater will tend to be the relative differences in imposition of the sanction and the smaller will tend to be relative differences in failure to impose the sanction. One cannot draw inferences about processes without being fully aware of these patterns, or, as is the near universal case, without even being aware that is possible for relative differences in an adverse outcome and relative differences in the corresponding favorable outcome to yield opposite conclusions about the comparative size of racial differences as to different types of outcomes.

**g. Illustration of patterns by which measures of differences between rates tend to be affected by the prevalence of an outcome with data on Black and white rates of one-or-more out-of-school suspensions and multiple out-of-school suspensions, with discussion of the implication of the patterns regarding the mistaken belief that generally reducing suspensions will tend to reduce relative racial differences in suspensions.**

Table 12 is based on data from OCR State and National Estimations files for the 2017-18 that allow one to calculate the rates at which Black and white students receive one-or-more out-of-school suspensions and rates at which they receive multiple suspensions. The table presents such rates with the same measures of difference presented in Table 3 through 11.

**Table 12. Black and white rates of one-or-more (OOM) out-of-school suspensions and multiple suspensions for the 2017-18 school year, with measures of difference.**

Suspension Frequency	Black Susp Rt	White Susp Rt	Bl/Wh Susp Ratio	Wh/Bl No Susp Ratio	Abs Diff (PP)	EES
OOM	12.01%	3.30%	3.64	1.10	8.72	.664
Multiple	5.21%	1.09%	4.78	1.04	4.12	.669

The table shows the usual pattern whereby each standard measure of difference between rates changes between the rows of information for the more common outcome of one-or-more suspensions and the less common outcome of multiple suspensions. Thus, whereas some would cite the relative differences in suspensions to say that the disparity is even worse for multiple suspensions than one-or-more suspensions, one could say the opposite based on the relative difference in rates of avoiding suspension and the absolute difference between rates. The EES figures are extremely close, indicating that the forces causing Black and white rates to differ are essentially the same with respect for one-or-more suspensions and multiple suspensions.

Most pertinent here, the table illustrate why providing an alternative to what would otherwise be a student’s first suspension will tend to increase relative racial differences in rates of one-or-more suspension. This is the opposite of what the DOE believes what it leads the public and other arms of government to believe.

**g. Illustration of patterns by which measures of differences between rates tend to be affected by the prevalence of an outcome with data on Black and white suspension rates in Oregon in the 2012/14 school year, with discussion regarding of its bearing on the mistaken belief that generally reducing suspensions will tend to relative racial differences in suspensions and activities of the PBIS community.**

Table 13 presents Black and white suspension rates in Oregon public schools in the 2007-08 and 2013-24 school years, with the same measures of difference in the previous table. The rates are from a presentation of Education Northwest titled “[Reducing Discipline Disparities – What Teachers are Saying](#)” given at the Northwest PBIS Network Spring Conference. A theme stressed in the presentation (at slide 5) was “Reduce the use of exclusionary discipline overall and to eliminate disparities.” Slide 7 showed an overall reduction in rates at which students received one or more suspensions from 7.1% in the 2007-08 school year to 3.3% in the 2013/14. Slide 8 showed that during that period the black rate decreased from 18.3% to 10.5% while the white rate decreased from 4.7% to 2.3%.

I do not know what was said about the differences in the Black and white rates might be measured. The title of the slide, however, is that “But, this is a common trend” suggests that the discussion involved the fact that the relative difference between Black and white rates increased. It is doubtful, however, that the discussion called into question the belief promoted by PBIS community that generally reducing suspensions would tend to reduce relative racial differences in suspensions.

**Table 13. Rates at which black and white students in Oregon were suspended one or more times in the 2000/08 and 2013/14 school years, with measures of difference**

School Year	Black Susp Rate	White Susp Rate	Blk/Wh Ratio Susp	Wh/Blk Ratio No Susp	Abs Diff (PP)	EES
2007-08	18.3%	4.7%	3.89	1.17	13.6	0.77
2013-14	10.5%	2.3%	4.57	1.09	8.2	0.74

In any case, the standard measures of differences between rates change in usual direction observed when there occurs a general reduction in suspension, while the EES suggest there was negligible change in the forces causing Black and white rates of being suspended or avoiding being suspended to differ.

As with Table 6 above, the illustration in the table differs little from similar illustrations that could be made from information pertaining to most states and countless local jurisdictions during a similar period of general reductions in suspension (including jurisdictions like Oakland and Denver where data like that in the table led school districts to mistakenly believe that restorative justice programs had reduced relative racial differences in suspensions). The illustration is presented here, however, because it the same illustration used in my July 5, 2015 [letter](#) to the first named author of the Girvan et al. article and other members of leadership of the University of Oregon Institute on Violence and Destructive Behavior and University of Oregon Law School Center for Dispute Resolution. The letter, which provided a specific example from Oregon, was a follow-up to a July 3, 2016 [letter](#) explaining both why one would expect to see such a pattern and that the pattern was in fact being observed all across country. Whether or not these letters or other communications to any of the authors influenced the Educational Psychologist article, the authors had ample reason at least to understand that the DOE and other agencies are mistaken in their view that generally reducing adverse discipline outcomes will tend to reduce relative racial differences in discipline rates.

Thus, the PBIS Center would seem to have an obligation to explain that issue to the DOE. This would so even if the Center believes there are reasons why PBIS programs will cause that not to happen or in some manner materially reduce the strength of the forces causing Black and white rates to differ. In any case, I suggest that it would behoove the agency for its leadership and technical staff to meet with representatives of the PBIS Center to discuss whether those representatives believe that generally reducing discipline rates will tend to increase or decrease relative racial and other differences in discipline rates and the proportion disadvantaged groups make up of disciplined students, as well as their reasons so believing. .

## **B. The Impossibility of Analyzing Demographic Differences by Comparing the Proportion a Group Makes Up of a Population with the Proportion It Makes Up of Persons Experiencing a Favorable or Adverse Outcome.**

This section discusses the impossibility of quantifying a demographic difference on the basis of a comparison between the proportion a group makes up of a population and the proportion it makes up of persons experiencing an outcome. An overriding point is that any presentation of data on group differences must present the actual rates at which the groups experience the outcomes.

Many equity issues are analyzed by comparing the proportion a group makes up of a population (PP) with the proportion it makes up of person experiencing a favorable outcome (PFO) or adverse outcome (PFO). This is often the approach for adverse criminal justice and school discipline outcomes, as well as many favorable educational outcomes, and has lately even been the approach with respect to COVID-19-related favorable or adverse health and healthcare outcomes. This is an unsound method of quantifying a difference for the same reasons that the

two relative differences and the absolute difference between rates are unsound measures of demographic differences. Specifically, PFO and PAO tend to be systematically affected by the prevalence of an outcome and hence any measure of difference between PP and PFO or PAO tends to be systematically affected by the prevalence of the outcome.

When there are only two groups in the population being examined, directions of change in the size of measures of difference between PP and PFO and between PP and PAO are the same as the directions of changes in relative differences in the associated favorable and adverse outcomes (though the matter becomes more complicated when the population is comprised of more than two groups<sup>33</sup>). For that reason alone, any difference between PP and either PFO or PAO would be an unsound measure of association for the same reasons that the two relative difference between rates are unsound measures of association.

Further, information on PP and PAO or PFO allows one to determine the relative differences between the rate at which the subject group experiences whichever of the two outcomes for which is information is presented and the rate at which all other persons experience the outcome.<sup>34</sup> But it does not enable one to determine the actual rates at which each group experiences the outcomes (and hence the relative differences in the corresponding adverse outcome or the absolute difference between rates, and, more important, a measure like EES). One must know the actual rates at which groups being compared experience an outcome in order to quantify the strength of the forces causing the favorable and corresponding adverse outcomes of the two group to differ or even to know the extent to which the proportion the group makes up of persons experiencing the outcome may be a function of the prevalence of an outcome.

But even persons who believe that either of the two relative differences (or the absolute difference) can effectively quantify the difference between the circumstances of the two group regarding the outcome and its opposite should recognize other reasons why comparing PP and PFO or PAO is an absurd way to quantify demographic differences. For one thing, when there are more than two groups in the population being examined, some groups may be found to be underrepresented among persons experiencing an adverse outcome or overrepresented among persons experiencing the favorable outcome, even though the groups' favorable and adverse outcome rates are worse than the rates for an advantaged group. For example, Hispanic students, whose rates for suspensions from school are usually greater than the rates for white students but lower than rates for Black students, are commonly regarded as underrepresented among suspended students even though their rates are higher than the rates of white students. And even when general reductions in suspensions are accompanied by increases in relative differences

---

<sup>33</sup> For this reason, I sometimes discuss the matter in terms of the way changing the prevalence of an outcome affected the proportion Blacks make up of the combined Black and white populations, as in Table 1 of "[Can We Actually Measure Health Disparities?](#)" *Chance* (Spring 2006).

<sup>34</sup> Knowing only that DG is 50% of the population (DGPP) and 65% of persons experiencing the adverse outcome of test failure (DGPAO) (as in Row 1 of Table 3), one can determine from the formula  $([DGPAO]/[DGPP])/((1-[DGPAO])/(1-[DGPP]))$  that the ratio of DG's failure rate to AG's failure rate is 1.85, the same ratio that would be calculated from the actual failure rates if they are known. But one cannot make sense out of the 1.85 ratio in the way a numerate observers might make sense out of the underlying failure rates (which also reveal the underlying pass rates).

between Hispanic and white suspension rates, in places where Black students make up a significant proportion of students, the proportion Hispanic students make up of suspended students may decrease (though the proportion Hispanic students make up of Hispanic and white students combined has increased).

Even when there are two only racial/ethnic groups in the population being examined, however, there are evident absurdities in quantifying demographic differences on the basis of comparison between PP and either PAO or PFO. One aspect of the problem is also pertinent to any comparisons between a group's outcome rate and the overall rate and may be more easily explained in that context. For the subject group's own rate influences the overall rate with which the group's rate is being compared. And the greater the proportion a group makes up of the population, the greater will be the influence of its rate on the overall rate, thus reducing all measures of difference between the group's rate and the overall rate. That is why it never makes sense to compare a disadvantaged group's rate with an overall rate rather than the rate of an advantaged group even when there are only two groups in the population.

Comparisons of PP with PAO or PFO are necessarily comparisons of the group's situation with an overall situation that is influenced by the group's own situation. But there are additional problems with comparing PP and PAO or PFO arising from the ways in which differences between and PAO or PFO are quantified. The May 2014 document "[Methods for Assessing Racial/Ethnic Disproportionality in Special Education](#)," which was funded by the DOE and produced by an arm of IDEA Data Center (a part of Westat, Inc.) recommends measuring differences between PP and PAO for assignment to special education or discipline in either relative or absolute terms. The April 2018 Government Accountability Office report titled "[Discipline Disparities for Black Students, Boys, and Students With Disabilities](#)" measured discipline disparities in terms of the absolute differences between the proportion Black students (or male students or students with disabilities) made up of students and the proportion they made up of students experiencing an adverse discipline outcome (PAO – PP), as did the DOE for certain matters in its April 2018 document "[Data Highlights on School Climate and Safety in Our Nation's Public Schools](#)." It used the same approach for quantifying demographic differences in rates of taking certain courses in the April 2018 OCR Report "[Data Highlights on Science, Technology, Engineering and Mathematics Course Taking in Our Nation's Public Schools](#)," though in that case employing the absolute difference to quantify the difference between the proportion a group makes up of students and the proportion it makes up of person experiencing a putatively favorable outcome.

In the DOE's June 2021 document "[An Overview of Exclusionary Discipline Practices in Public Schools for the 2017-2018 School Year](#)" and in the Discipline RFI the agency measures discipline disparities in terms of relative differences between the proportion a group or subgroup makes up of students and the proportion it makes up of students experiencing an adverse discipline outcome – that is,  $(PAO - PP)/PP$  or  $(PAO/PP) - 1$ , while sometimes characterizing the matter in terms of the ratio of PAO to PP.

At the same that an increase in the proportion a group makes up of persons experiencing the outcome affects the situation of the overall population with whose situation the group's

situation is being compared, the increase also affects relative differences and absolute differences between PP and PAO, and it does so in conflicting ways, as illustrated in Table 14.

**Table 14. Relative and absolute differences between proportion Black students make up of all students and proportion Black students make up of suspended students in schools where Black and white suspension rates are 15% and 5% with various Black proportions of all students.**

Black Rate	Black Prop Students	Black Prop Suspensions	Abs Df (PP)	Rel Df
15%	5%	13.64%	8.64	172.73%
15%	20%	42.86%	22.86	114.29%
15%	40%	66.67%	26.67	66.67%
15%	60%	81.82%	21.82	36.36%
15%	80%	92.31%	12.31	15.38%
15%	95%	98.28%	3.28	3.45%

Table 14 presents situations where all students are either Black or white and Black and white suspension rates are 15% and 5%, though the proportion Black students make up of students varies. There is no basis for distinguishing between schools that have those rates with the respect to the magnitude of the difference (though one might find varying reasons why Black and white rates differ in each situation). Consider what the table illustrates regarding the way the proportion Black students make up of all students affects absolute and relative differences between PP and PAO when the Black proportion of students increases incrementally from 5% to 95%.

The relative difference between PP and PAO decreases consistently from the point where Black students make up 5% of students (and PAO is 172.7% greater than PP) to the point where Black students make up 95% of students (and PAO is only 3.24% greater than PP).<sup>35</sup> The absolute difference between PP and PAO, however, is affected by increases in PP in a more complicated way. It is 8.64 percentage points when Black students make up only 5% of students, increases until reaching a maximum of 26.7 percentage points when Black students make up approximately 40% of students, and then declines until reaching 3.28 percentage points when Black students make up 95% of students.

Not only do both approaches find differences in disparities from school-to-school even though the situation is exactly the same in each school, the two approaches can yield different views as to which of two schools has the larger disparity problem. Notice that in the schools

---

<sup>35</sup> As shown in Table 7 of my CCR testimony (at Loudoun County Public schools has quite low suspension rates for both Black and white student compared with national averages. It also has a very low EES for the difference between Black and white rates. While the comparatively low suspension rates drives up the relative racial differences in suspension rates, the very low EES causes it to remain slightly below the national average. But as discussed on the [Loudoun County \(VA\) Disparities](#), because Black students made up a comparatively low proportion of students in the district, the school district, which measured disparities in terms of relative difference between the proportion Black students made up of students and the proportion they made up of suspended students, the district had believed it had an especially larger racial disparity in suspensions.

where Black students make up 20% and 40% of students, the former school shows the larger relative difference while the latter school shows that larger absolute difference.

The same issues exist when racial difference in favorable outcomes are examined by comparing the proportion Black students make up of students (PP) with the proportion they make up of students experiencing a favorable outcome (PFO) like assignment to a gifted and talented program. Table 15 presents a situation where gifted and talented rates are 5% for Black students and 15% for white students in each school being examined but where the proportion Black students make up of students varies in the preceding table. The negative signs in the final two columns reflect the fact that the values show the degree to which the Black PFO is less than the proportion Black students make up of all students.

**Table 15. Relative and absolute differences between proportion Black students make up of students and proportion Black students make up of students in gifted and talented programs in schools where Black and white rates of assignment to programs are 5% and 15% with various Black proportions of all students.**

<b>Black Rate</b>	<b>White Rate</b>	<b>Bl Prop Students</b>	<b>Bl Prop GIFT</b>	<b>Ab Df (PP)</b>	<b>Rel Df</b>
5%	15%	5%	1.72%	-3.28	-65.52%
5%	15%	20%	7.69%	-12.31	-61.54%
5%	15%	40%	18.18%	-21.82	-54.55%
5%	15%	60%	33.33%	-26.67	-44.44%
5%	15%	80%	57.14%	-22.86	-28.57%
5%	15%	95%	86.36%	-8.64	-9.09%

The table shows patterns similar to those in Table 7, though the absolute and relative difference values are somewhat different from those in Table 7. Also, Black students are a larger proportion of students (approximately 63%) at the point where the percentage point difference reaches a maximum (in negative terms). As in the prior example respecting schools with 20% and 40% of students, the former school shows the larger relative difference while the latter school shows the larger absolute difference.

Now consider how this approach might cause observers to find suspension differences to be larger in one school than another when they are in fact smaller. School A has the same Black and white 15% and 5% rates suspension rate as those in the earlier example, while School B has Black and white suspension rates of 13% and 5%. Because the white rate is the same in the two situations all measures of difference between Black and white rates would be larger in School A than in School B.

In this situation, as shown in Table 16, even though the racial difference is greater in School A than School B, if Black students were 10% of students in School A and 20% of students in School B, the absolute difference between PP and PAO would be greater in School B than School A (19.39 versus 15.0 percentage points). On the other hand, if Black students were 20% of students in School A and 10% of students in School B, the relative difference between PP and PAO would be greater in School B than School A (124.14% versus 114.97%).

**Table 16. Relative and absolute differences between proportion Black students make up of all students and proportion Black students make up of suspended students in schools where Black and white suspension rate are 15% and 5% (School A) or 13% and 5% (School B) with various Black proportions of all students of 10% and 20%.**

School	Black Rate	White Rate	Black Prop Students	Black Prop Suspensions	Abs Df (PP)	Rel Df
A	15%	5%	10%	25.00%	15.00	150.00%
A	15%	5%	20%	42.86%	22.86	114.29%
B	13%	5%	10%	22.41%	12.41	124.14%
B	13%	5%	20%	39.39%	19.39	96.97%

As discussed in Section A, the RFI (and many recent DOE documents) attempt to quantify disparities by means of a comparison between the proportion certain racial/gender subgroups group makes up of students and the proportion the subgroups makes up of students experiencing an outcome. While I have not examined how the issues illustrated above would affect such comparison, I assume that they would be similar though somewhat more complicated.

For fuller discussions of the problems with this attempting to analyze a demographic difference on the basis of a comparison between the proportion a group makes up of a population and the proportion it makes up of persons experiencing an adverse or favorable outcome, see Kansas Law Paper (at 23-26), CEP Comments (at 43-45), TDHCD Brief (at 23-27), [IDEA Data Center Disproportionality Guide](#) subpage of the [Discipline Disparities](#) page, and slides 97-108 of the University of Maryland [workshop](#). But even the discussion in this section should make abundantly clear that the DOE does the public a great disservice by presenting proportions a group make up of students and the proportion the group of students experiencing a favorable or adverse outcome rather than the actual rates at which members of each disadvantaged group and corresponding advantaged comparison group experiences whatever favorable or adverse outcome is at issue. That is so both because it never makes sense to compare a group's situation with the situation of the overall population whose situation is influenced by the group's situation and because it is only possible to effectively quantify demographic differences based on the actual rates at which advantaged and disadvantaged groups experience the outcome.