

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

August 25, 2017

HAND DELIVERED

David Myers, President and CEO
Patricia Y. Gurin, Chair of the Board of Directors
American Institutes for Research
1000 Thomas Jefferson Place
Washington, DC 20007

Re: Problems in AIR Analyses of Demographic Differences Involving
Outcome Rates Resulting From Failure to Recognize Patterns by Which
Measures Tend to Be Affected by the Prevalence of an Outcome

Dear Dr. Myers and Professor Gurin:

This letter has three purposes. The first purpose is to explain to the American Institutes for Research (AIR) the ways that its analyses of demographic differences involving favorable and adverse outcomes are undermined by a failure to understand the ways the measures employed in such analyses tend to be affected by the prevalence of an outcome. A second purpose is to alert AIR of the extent to which its funding by the Department of Education and other agencies of the federal government is in jeopardy as a result that failure of understanding. A third purpose is to advise AIR of its responsibilities to explain to the Department of Education and other agencies with which it has contracts of the ways the activities of those agencies and their contractors and grantees are undermined as a result of that failure of understanding.

Attached is a July 17, 2017 letter to heads of the U.S. Departments of Education (DOE), Health and Human Services (HHS), and Justice (DOJ). The letter is also available [online](#),¹ as are earlier letters to DOE or DOE and HHS dated [August 24, 2015](#), and [April 18, 2012](#), and earlier letters to DOJ or DOJ and the City of Ferguson, Missouri dated [April 13, 2017](#), [March 9, 2015](#), and [April 23, 2012](#). The July 17, 2017 letter is discussed in my "[Innumeracy at the Department of Education and the Congressional Committees Overseeing It](#)," Federalist Society Blog (Aug. 24,

¹ To facilitate consideration of issues raised in documents such as this I include links to referenced materials in electronic copies of the documents, in some cases, for the reader's convenience, providing the links more than once. Such copies are available by means of the [Measurement Letters](#) page of jpscanlan.com. If the online version of the letter is amended, such fact will be noted on the first page.

2017), and it and the April 13, 2017 letter to DOJ are discussed in my "[The Government's Uncertain Path to Numeracy](#)," Federalist Society Blog (July 21, 2017).

The principal subject of the July 17 letter involves the fact the three recipient agencies, like many social scientists and research institutes, have for some time led the public and school administrators to believe that generally reducing discipline rate rates will tend to reduce both (a) relative (percentage) racial and other differences in discipline rates and (b) and the proportions racial minorities and other more susceptible groups make up of disciplined students. The letter explains that, in fact, generally reducing discipline rates tends to increase, not reduce, both (a) and (b). That is, while generally reducing discipline rates will tend to reduce relative differences in rates of avoiding discipline, it will tend to increase relative differences in discipline rates; while generally reducing discipline rates will tend to increase the proportions groups more susceptible to discipline make up of students avoiding discipline, it also will tend to increase the proportions such groups make up of students who are disciplined.

The letter (at 9 n.8) references the [Restraint Disparities](#) subpage of the [Discipline Disparities](#) page of [jpscanlan.com](#), which discusses the failure of DOE to understand that the more school administrators follow DOE guidance to limit the physical restraint of students, the greater will tend to be the proportion students with disabilities make up of restrained students. That page references a 2015 DOE-funded AIR study titled "[Addressing the Root Causes of School Discipline – An Educator's Action Planning Guide](#)," which I will discuss further below.

The letter to the three agency heads urges them at least to explain to school administrators that the agencies' prior guidance regarding the effects of relaxing standards and otherwise generally reducing discipline rates on measures of demographic differences in discipline outcomes was incorrect.

But the letter also discusses that essentially all federally-funded research into demographic differences in outcome rates, including that conducted by federal agencies themselves and that conducted by recipients of federal funds, has been undermined by a failure to recognize patterns by which measures of differences between outcome rates tend to be affected by the prevalence of an outcome. These patterns include that whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it, which the National Center for Health Statistics recognized more than a decade ago. They also include the more complicated patterns by which the absolute difference between rates and the difference measured by the odds ratio tend to be affected by the prevalence of an outcome.²

² Roughly, as an outcome goes from being rare to being fairly common, absolute differences tend to increase; as the outcome goes from being fairly common to being even more common, absolute differences tend to decrease. As the prevalence of an outcome changes, the difference measured by the odds ratio tends to change in the opposite direction of the absolute difference. While all measures may change in the same direction, anytime a relative difference and the absolute difference change in opposite directions, the other relative difference will necessarily have changed in the opposite direction of the first relative difference the same direction as the absolute difference.

The letter therefore recommended that the agencies undertake a complete review of the soundness of the methods by which they have analyze demographic differences and of the soundness of the guidance they have provided on such analyses. It also recommends that the agencies institute moratoriums on grants and contracts (and activities pursuant to grants and contracts already awarded) to which the measurement issues discussed in the letter pertain. See especially the discussion at page 3 of the letter regarding DOE funding.

I am not familiar with all AIR research on demographic differences, even that funded solely by the Department of Education. But my familiarity with some of the organization's DOE-funded research is such that I believe that the measurement issues discussed in the July 17 letter likely pertain to all DOE-funded AIR research on demographic differences involving educational outcomes. Such issues also pertain to the organization's DOE-funded research involving school environments, which commonly is premised on a mistaken belief as to the effects of generally reducing discipline rates on the measures of demographic differences involving discipline that DOE principally employs. Presumably, the issues addressed in the letter also pertain to most or all AIR research into demographic differences funded by other federal agencies, including that pertaining to health and healthcare disparities.

The key elements of the mistaken understanding of the effects of generally reducing discipline rates on standard measures of discipline disparities are illustrated in the tables of the attached letter. They are also treated fairly succinctly in my "[Things government doesn't know about racial disparities](#)," *The Hill* (Jan. 28, 2014), "[The Paradox of Lowering Standards](#)," *Baltimore Sun* (Aug. 5, 2013), and "[Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies](#)," *Amstat News* (Dec. 2012).

A more extended treatment of this mistaken understanding and the broader problems with analyses of demographic differences that fail to consider the ways the measures employed in the analyses tend to be affected by the prevalence of an outcome may be found in my "[Race and Mortality Revisited](#)," *Society* (July/Aug. 2014). That article shows that the failures of understanding that I attribute to AIR research above or below are essentially universal among persons analyzing demographic differences in the law and the social and medical sciences.

Other extended treatments may be found in my "[The Mismeasure of Health Disparities](#)," *Journal of Public Health Management and Practice* (July/Aug. 2016), [Comments for the Commission on Evidence-Based Policymaking](#) (Nov. 14, 2016),³ [Letter to Population Association of America and Association of Population Centers](#) (Mar. 29, 2016), [Letter to American Statistical](#)

³ See especially the recommendations at pages 47-48 regarding the funding issues addressed in this letter. Pages 3-4 of follow-up [comments](#) dated November 28, 2016, address funding issues with regard to the mistaken belief that Positive Behavioral Interventions and Support programs will tend to reduce relative differences in discipline rates and the proportions groups more susceptible to discipline make up of disciplined students..

[Association](#) (Oct. 8, 2015), "[Measuring Health and Healthcare Disparities](#)," Proceedings of the Federal Committee on Statistical Methodology 2013 Research Conference (Mar. 2014), "[The Mismeasure of Discrimination](#)," Faculty Workshop, University of Kansas School of Law (Sept. 20, 2013), [Letter to Harvard University](#) (Oct. 9, 2012), "[Can We Actually Measure Health Disparities?](#)," *Chance* (Spring 2006), and "[Race and Mortality](#)," *Society* (Jan./Feb. 2000).

Most of the items listed in the paragraph above provide some graphical or tabular illustrations of the pertinent statistical patterns and implications of the failures to understand them. Many more such illustrations may be found in the following methods workshops at American educational institutions: "[The Mismeasure of Health Disparities in Massachusetts and Less Affluent Places](#)," Department of Quantitative Health Sciences, University of Massachusetts Medical School (Nov. 18, 2015); "[The Mismeasure of Discrimination](#)," Center for Demographic and Social Analysis, University of California, Irvine (Jan. 20, 2015); "[The Mismeasure of Demographic Differences in Outcome Rates](#)" Public Sociology Association of George Mason University (Oct. 18, 2014); "[Rethinking the Measurement of Demographic Differences in Outcome Rates](#)," Maryland Population Research Center of the University of Maryland (Oct. 10, 2014); "[The Mismeasure of Association: The Unsoundness of the Rate Ratio and Other Measures That Are Affected by the Prevalence of an Outcome](#)," Minnesota Population Center and Division of Epidemiology and Community Health of the School of Public Health of the University of Minnesota (Sept. 5, 2014); "[The Mismeasure of Group Differences in the Law and the Social and Medical Sciences](#)," Institute for Quantitative Social Science at Harvard University (Oct. 17, 2012); "[The Mismeasure of Group Differences in the Law and the Social and Medical Sciences](#)," Department of Mathematics and Statistics of American University (Sept. 25, 2012). Conference presentations employing similar illustrations, sometimes based on data pertaining to the country where the conference took place, may be found [here](#).

See also the scores of web pages on [jpscanlan.com](#) addressing these issues, of which the most pertinent to AIR federally-funded research are the [Discipline Disparities](#), [Educational Disparities](#), [Measuring Health Disparities](#), [Immunization Disparities](#), and [Mortality and Survival](#) pages and the subpages to the first three of these pages. Many of the letters collected on the [Measurement Letters](#) page explain issues pertaining to the recipients' activities that would apply as well to the activities of AIR.

With regard to the importance of AIR researchers' understanding the issues addressed in the above materials, I call your particular attention to the discussion of Table 2 of "Race and Mortality Revisited" (at 329-330,343) concerning whether there can be justification for devoting resources to studying the role of particular policies with respect to an observed change in any measure of racial differences in poverty (or its avoidance) without consideration of the implications of the way the measures examined tend to be affected by general changes in the level of poverty. But I urge you also to consider the same point with regard to the patterns shown in each of the tables in the article.

Further, Table 2 of "Race and Mortality Revisited" shows how general reductions in poverty will tend to increase relative racial differences in poverty rates while reducing relative racial differences in rates of avoiding poverty. It also shows that, given the rates at which whites and African Americans in the United States commonly experience poverty, a decline in poverty will tend to reduce absolute differences between the two groups' poverty rates. The rate ranges at issue for public school suspensions are commonly such that general reductions in discipline rates will tend to result in changes on measures of racial disparity similar to those observed in the case of a general reduction in poverty. That is, relative differences in suspension rates will tend to increase while relative differences in rates of avoiding suspensions will tend to decrease (a pattern that will typically be observed regardless of the rate ranges at issue), and absolute differences between rates will tend to decrease (a pattern contingent on the rate ranges at issue).

The aforementioned 2015 AIR document titled "[Addressing the Root Causes of School Discipline – An Educator's Action Planning Guide](#)," while frequently mentioning "overrepresentation" of certain groups among students disciplined and "disproportionality" (or some variation on the word) in discipline, discusses (at 34-35) measurement of discipline disparities in terms of both relative differences between discipline rates and absolute differences between discipline rates. And it does mention the possibility that the relative difference and the absolute difference can change in different directions. But the document contains no suggestion of an understanding of the reasons to expect this to occur systematically, or that the proportion an overrepresented group makes up of students experiencing an outcome (which is associated with the relative difference for the outcome⁴) will tend to increase when the absolute difference decreases. The document also contains no suggestion of an understanding of the way general reductions in suspension rates will tend to affect each measure of disparity, including that such reductions will tend to increase overrepresentation and disproportionality (as these terms are most commonly used). Most important, given the document's emphasis on understanding causes of discipline disparities, the document provides no guidance for divining what light changes in

⁴ While I often describe the pattern whereby the rarer an outcome the greater tends to be the proportion the more susceptible group makes up of persons experiencing the outcome as a corollary to the pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it (or vice-versa) (see, e.g., page 9 of the October 8, 2016 letter to the American Statistical Association), the correlation is exact only when there are just two groups. Commonly, issues are discussed in terms of the proportion a group makes up of all persons experiencing an outcome (which is related to the relative difference between the rate at which the group experiences and the rate at which all other persons experience the outcome.) Thus, for example, when suspension rates decline, even when the relative difference between rates for whites and the rates for each disadvantaged racial/ethnic minority increases, the proportion some such groups make up of all persons suspended may decline. And often, even though a disadvantaged racial/ethnic minority may have a higher suspension rate than whites, the group may be deemed to be underrepresented among persons suspended. These are among the many reasons that a disparities issue can never be soundly analyzed in terms of the difference between the proportion a group makes up of persons potentially experiencing an outcome the proportion it makes up of persons actually experiencing the outcome. See the [IDEA Data Center Disproportionality Guide](#) subpage of the [Discipline Disparities](#) page of [jpscanlan.com](#), Section C of the Kansas Law paper titled "The Mismeasure of Discrimination" and Section C of the November 14, 2016 comments for the Commission on Evidence-Based Policymaking.

measures shed on whether the factors causing the rates of advantaged and disadvantaged groups to differ are growing stronger or weaker over time.⁵

More broadly with respect to the importance that AIR address with its researchers the issues addressed in the references above, the organization should consider the following:

1. Probably no AIR document has reflected an understanding that measures tend to change solely because the prevalence of an outcome changes.
2. Possibly no researcher at AIR is aware 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 outcome changes, much less that more than a decade ago the National Center for Health Statistics recognized that such pattern tends to occur systematically.
3. Probably most and possibly all AIR researchers believe that reducing adverse school discipline outcomes or other adverse outcomes should tend to reduce relative differences in rates of experiencing the outcomes.

As previously suggested, and as reflected in materials referenced above,⁶ these failures of understanding on the part of AIR researchers are no different from the failures of understanding of persons analyzing demographic differences at other research institutions, including those commonly deemed the most prestigious in the world. But that is not justification for AIR to fail to ensure that its researchers understand these issues or justification for AIR's securing funding from the government or other entities to produce research that fails to reflect an understanding of the issues. And, as suggested in the recent posts cited at the outset (see also my "[Will Trump Have the First Numerate Administration?](#)," Federalist Society Blog (Jan. 4, 2017)), there is some prospect, if still an uncertain one, that the government may soon recognize the wastefulness of research into demographic differences that fails to understand implications of the patterns by which measures tend to be affected by the prevalence of an outcome. Thus, AIR should want to

⁵ See the discussion in "Race and Mortality Revisited" of its Table 5. See also the discussion at page 12 of the April 13, 2017 [letter](#) to DOJ regarding the absence of any basis for distinguishing between the two rows of the letter's Table 1 (which is also Table 1 of the attached July 17, 2017 letter to the three agencies) with regard to the likelihood that the patterns result from discrimination. See also the September 20, 2016 [letter](#) to Oklahoma City School District regarding the difficulties of complying with an agreement with the Department of Education that envisions both general reduction in discipline rates and reduction in the proportions African Americans make up of disciplined students. See also the September 12, 2016 [letter](#) to Antioch Unified School District regarding the problems facing that district as a result of the failure to understand the effects of modifications of practices on different measures of disparity. See generally aforementioned Kansas Law paper titled "[The Mismeasure of Discrimination](#)" regarding the impossibility of appraising the likelihood that a difference between rates indicates bias without understand the way the measure employed tends to be affected by the prevalence of an outcome.

⁶ See letters to institutions and organizations collected in the [Measurement Letters](#) page of [jpscanlan.com](#).

ensure that its current and future research can be fully defended. AIR should also want to be in position to assist the government in addressing the longstanding failures of understanding in government-conducted or government-funded analyses of demographic differences involving outcome rates.

Turning to the third purpose of this letter, I note that at pages 4-5 of the November 14, 2016 [comments](#) for the Commission on Evidence-Based Policymaking, I discuss the reasons proffered by the American Statistical Association (ASA) and the Population Association of America (PAA) for refusing my requests (a) that they form committees to address problems in standard analyses of demographic differences and (b) that they advise the government of its mistaken understanding regarding the effects of reducing the frequency of adverse outcomes on standard measures of racial and other demographic differences regarding the outcomes. I suggest that the refusals do not reflect responsible actions of scientific organizations, particularly when the organizations commonly write to Congress to support funding of the research into demographic differences.

But whether or not the refusals of ASA and PAA can be justified, AIR is in a position that is substantially different from that of ASA or PAA. For AIR receives substantial funds from DOE and other agencies for the purpose of promoting the soundness of statistical analyses pertaining to the activities of such agencies. Acceptance of such funds imposes an obligation on the organization to advise the agencies of ways in which statistical analyses pertaining to their activities are unsound, including the ways AIR's own federally-funded research has been unsound. An obligation to explain to the agencies that their understandings of the effects of reducing adverse outcomes on measures of demographic difference regarding the outcomes is the opposite of reality is but one element of the larger obligation. But, given the many perverse consequences of policies based on that understanding (as explained in the three short items mentioned at page 3 above and the letter to Oklahoma City School District discussed in note 5⁷), it is a pressing element of the obligation. Thus, it is a matter to which AIR should give attention as soon as possible.

This item is addressed only to AIR President and COE David Myers and AIR Board Chair Patricia Y. Gurin. But I am circulating it by email to other members of AIR leadership and certain staff members for whom I could find email addresses. And I urge you to circulate it widely within the organization as well. For, even if AIR leadership were of the view that the issues I raise merely involve a difference of scientific opinion from that underlying AIR

⁷ See also "[The Perverse Enforcement of Fair Lending Laws](#)," *Mortgage Banking* (May 2014) and my *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). See Tables 2 and 3 of the April 13, 2017 [letter](#) to DOJ for illustrations of the way that lowering income or credit score requirements to secure loans or other favorable outcomes, while tending to reduce relative racial differences in meeting the requirements, will tend to increase relative racial differences in failing to meet the requirements.

David Myers, President and CEO
Patricia Y. Gurin, Chair of the Board of Directors
August 25, 2017
Page 8

research,⁸ AIR researchers deserve to be made aware of the issues. They also deserve to be made aware that the commonplace or nearly universal belief that reducing the prevalence of outcome will tend to reduce relative differences in rates of experiencing the outcome is incorrect.

Sincerely

/s/ James P. Scanlan

James P. Scanlan

Attachment

⁸ See "Race and Mortality Revisited" at 344 regarding the reasons proffered by Harvard Medical School and Massachusetts General Hospital for refusing to withdraw a healthcare disparities measurement guide that failed to show any understanding of patterns by which the measures it discussed tended to be affected by the prevalence of an outcome. See my October 26, 2012 [letter](#) to the entities responsible for the guide.

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

July 17, 2017

The Honorable Betsy DeVos
Secretary of Education
United States Department of Education
400 Maryland Avenue, SW
Washington, DC 20202

The Honorable Thomas E. Price, M.D.
Secretary of Health and Human Services
United States Department of Health and Human Services
200 Independence Avenue, SW
Washington, DC 20201

The Honorable Jeff Sessions
Attorney General
United States Department of Justice
950 Pennsylvania Avenue, NW
Washington, DC 20530-0001

Re: Obligations of the Departments of Education, Health and Human Services, and Justice to Correct Their Erroneous Guidance Suggesting That Relaxing Discipline Standards Tends to Reduce, Rather Than Increase, (a) Relative Demographic Differences in Discipline Rates and (b) the Proportions Groups More Susceptible to Adverse Discipline Outcomes Make Up of Persons Experiencing the Outcomes

Dear Secretary DeVos, Secretary Price, and Attorney General Sessions:

The purpose of this letter to advise the Departments of Education (DOE), Health and Human Services (HHS), and Justice (DOJ) of an obligation to correct erroneous guidance the three agencies have been providing the public, policymakers, and school administrators regarding the relationship between the stringency of school discipline standards and racial and other demographic differences in discipline outcomes. At least since the early years of this decade DOE and DOJ have been promoting the belief that relaxing standards and otherwise reducing rates of suspension and other adverse discipline outcomes will tend to reduce (a) relative (percentage) racial and other demographic differences in rates of experiencing the outcomes and (b) the proportions more susceptible groups make up of persons experiencing the outcomes. In

December 2014, the Secretary of HHS, in a document titled “[Policy Statement on Expulsion and Suspension Policies in Early Childhood Settings](#)” (Policy Statement) and an associated [Dear Colleague Letter](#), joined the Secretary of Education in promoting the belief that generally reducing adverse discipline outcomes would tend to reduce (a) and (b).

In fact, generally reducing any outcome tends to increase both (a) and (b) as to the outcome. Thus, the agencies have been leading a wide range of persons and entities to believe something about an important matter that is the opposite of reality. In any situation where government agencies have provided misleading guidance to the public the agencies have an obligation to correct the misleading guidance. The obligation is heightened where, as here, the agencies represent themselves to have, or are assumed by the public to have, expertise in the matter.

I briefly explain below the pertinent statistical point, which I have recently also explained in an April 13, 2017 [letter](#)¹ to Attorney General Jeff Sessions and Acting Assistant Attorney General T. E. Wheeler, III (Sessions letter) and in other communications to DOJ attorneys. Before doing so, however, I make certain preliminary points regarding the relationship of the principal subject of this letter to larger subjects the agencies must address if they are to fulfill their missions in a responsible manner.

Preliminary points regarding the instant subject and the larger subjects the agencies must address

This letter focuses on a discrete matter that agency officials, once having focused on a statistical pattern recognized more than a decade ago by the National Center for Health Statistics, should understand both to be undebatable and to involve agency actions that are the antithesis of responsible government. Further, the matter is something the three agencies can immediately begin to address at least by a Dear Colleague Letter explaining that express or implied guidance in prior such letters was incorrect. The matter also is quite pressing because thousands of school administrators across the country are continually endeavoring to implement policies promoted by the government (or incorporated into agreements with the government) while relying on the government’s mistaken guidance as to the effects of those policies on the measures of demographic differences that the government employs.² Numerous state and local governmental authorities have already taken actions based on the government’s erroneous guidance and others are considering like actions.

¹ To facilitate consideration of issues raised in documents such as this I include links to referenced materials in electronic copies of the documents, in some cases, for the reader’s convenience, providing the links more than once. Such copies are available by means of the [Measurement Letters](#) page of [jpscanlan.com](#). If the online version of the letter is amended, such fact will be noted on the first page.

² The matters is particularly pressing in the case of the school districts acting pursuant to agreements with DOE where the agency’s failure of understanding has created situations in which the more the school districts (or parts thereof) endeavor to comply with the agreement the more likely it is that DOE will regard them to have violated the agreement. See my September 20, 2016 [letter](#) to Oklahoma City School District.

But the agencies should recognize that the failure of understanding of elementary statistics that has led the agencies to provide the aforementioned mistaken guidance is part of a larger failure of understanding on the part of the agencies regarding the ways measures commonly employed in the analyses of demographic differences tend to be affected by the prevalence of an outcome. As a result of the larger failure of understanding, virtually nothing the agencies have themselves done, or that has been done pursuant to grants and contracts awarded by the agencies, regarding the analyses of demographic differences involving outcome rates has been statistically sound. See, *e.g.*, my [Comments for Commission on Evidence-Based Policymaking](#) (Nov. 14, 2016) (first CEBP comments), “[The Mismeasure of Health Disparities](#),” *Journal of Public Health Management and Practice* (July/Aug. 2016), “[Race and Mortality Revisited](#),” *Society* (July/Aug. 2014), and “[Measuring Health and Healthcare Disparities](#),” Proceedings of Federal Committee on Statistical Methodology 2013 Research Conference (March 2014). See also my “[Will Trump Have the First Numerate Administration?](#)” Federalist Society Blog (Jan. 4, 2017), regarding prospects that the current administration will be able understand things about analyses of demographic differences that other administrations have failed to understand.

In the case of DOE, the larger failure of understanding has prevented the agency from conducting any useful analyses of whether racial differences in educational outcomes like retention in grade, graduation, proficiency, assignment to disabled status, and various other matters have increased or decreased over time. See the [Educational Disparities](#) page of [jpscanlan.com](#) and its subpages, my August 24, 2015 [letter](#) to the HHS Secretary Sylvia M. Burwell and DOE Secretary Arne Duncan (at 9-11), and my April 18, 2012 [letter](#) to DOE Secretary Arne Duncan and Assistant Secretary of Education for Civil Rights Russlyn Ali (at 4). For example, as proficiency rates generally improve, relative demographic differences in rates of achieving proficiency tend to decrease while relative differences in rates of failing to achieve proficiency tend to increase; as proficiency rates generally improve, absolute demographic differences between rates of achieving basic proficiency (where rates are often well above 50 percent) tend to decrease, while absolute differences between rates of achieving advanced proficiency (where rates usually are well below 50 percent) tend to increase.³ To my knowledge, nothing DOE or any entity assisting it has done regarding analyses of demographic differences involving outcome rates has reflected an awareness of these patterns. Thus, DOE should undertake a complete review of the soundness of the methods by which it has analyzed demographic differences and of the soundness of the guidance it has provided on this subject. The agency should also institute a moratorium on grants and contracts (and activities pursuant to grants and contracts already awarded) to which these measurement issues pertain.⁴

³ Examples of these patterns may be found in the [Education Trust Glass Ceiling Study](#) subpage of the [Educational Disparities](#) page of [jpscanlan.com](#).

⁴ A minimum requirement of federally-funded research on demographic differences in outcome rates should be a commitment of the researchers to attempt to address the implications of the effects of the frequency of an outcome on the measures employed in the research. See fourth recommendation of the [first CEBP comments](#) (at 47). But the measurement issues addressed in those comment are pertinent both to activities involving analyses of demographic differences and activities that, while not necessarily involving analyses of such differences, are based on mistaken understandings regarding effects of policies on measures of demographic differences. The latter include, for example, activities that are based on the mistaken belief that positive behavioral intervention and support programs will tend to reduce relative racial differences in discipline rates, as in the case of the \$1 million grant discussed in

In the case of HHS, as discussed in the references at the top of page 3, the larger failure of understanding has led to the expenditure of many billions of dollars in research into demographic differences in health and healthcare outcome that has yielded very little of value even when it has not been patently misleading. One of the many situations exemplary of the failures of understanding on the part of HHS and its arms is the following. The National Center for Health Statistics (NCHS) more than a decade ago recognized that, as health and healthcare improve relative differences in favorable health and healthcare outcomes and relative differences in the corresponding adverse outcomes tend to change systematically in opposite directions as the prevalence of an outcome changes; yet, so far as the published record reveals, no other arm of HHS has recognized that it is even possible for relative differences in a favorable health and healthcare outcome and relative differences in the corresponding adverse outcome to change in opposite directions as the prevalence of an outcome changes. To my knowledge, no health or healthcare disparities research conducted or funded by arms of HHS has considered whether an observed pattern of changes in a measure employed in the research was anything other than a function of the change in the prevalence of the outcome. See the first four references at the top of page 3 and my [“The Mismeasure of Health Disparities in Massachusetts and Less Affluent Places,”](#) Quantitative Methods Seminar, Department of Quantitative Health Sciences, University of Massachusetts Medical School (Nov. 18, 2015). The points in the last two sentences of the prior paragraph regarding DOE apply equally to HHS.

In the case of DOJ, the consequences of the larger failure of understanding are summarized to a degree in the Sessions letter and include many situations where the more an entity complies with DOJ guidance (or obligations imposed by decrees in suits brought by the DOJ) the more likely the entity is to be sued by DOJ (or found not to comply with decree-imposed obligations). See my [“Compliance Nightmare Looms for Baltimore Police Department,”](#) Federalist Society Blog (Feb. 8, 2017), [“Things DoJ doesn’t know about racial disparities in Ferguson,”](#) *The Hill* (Feb. 22, 2016), [“Things government doesn’t know about racial disparities,”](#) *The Hill* (Jan. 28, 2014), [“Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies,”](#) *Amstat News* (Dec. 2012). See also my [Comments on the Selection of Monitor of the Baltimore Police Consent Decree](#) (June 26, 2017) regarding the unlikelihood that the experts identified in the monitor proposals for the consent decree covering Baltimore Police practices understand the effects of reducing adverse criminal justice on measures of demographic differences any better than the government does.

Thus, each of the agencies has a responsibility to examine the problems in the analyses of demographic differences that it conducts or funds with an aim toward ensuring that future analyses are sound and that no further research, even on existing grants and contracts, continues to employ unsound methods. I may contact the agencies again regarding such matters. But there is no need for the agencies to await such contacts before examining the extent to which their failures to understand the ways measures tend to be affected by the prevalence of an outcome have undermined their activities.

Further, Section 5 of the Evidence-Policymaking Commission Act of 2016 imposes on each of the heads of DOE, HHS, and DOJ a responsibility to advise and consult with the Commission on Evidence-Based Policymaking regarding matters within the agency heads' areas of responsibility. Thus, the aforementioned reviews by DOE and HHS (and like actions suggested in the Sessions letter) should be conducted in a sufficiently timely fashion for the agencies to fulfill their responsibility to the Commission on Evidence-Based Policymaking before the Commission issues its report to Congress and the President this fall. I suggest that my comments for the Commission dated [November 14, 2016](#), and [November 28, 2016](#), provide the agencies a useful guide for advising the Commission as to the ways the agencies' missions have so far been undermined by the failure to understand the statistical patterns described in the comments.

Attention to these larger subjects, however, should not interfere with the agencies' fulfilling their responsibilities to immediately correct their guidance regarding the effects of relaxing discipline standards on measures of difference in school discipline outcomes.

Patterns by which restricting adverse outcomes to those most susceptible to them tends to increase measures of demographic differences as to the outcomes

For reasons related to the shapes of underlying distributions of factors associated with experiencing an outcome or its opposite, all standard measures of differences between outcome rates (*i.e.*, the proportions of demographic groups experiencing a binary outcome) tend to be affected by the frequency of an outcome. The pattern most pertinent here is that whereby the rarer an outcome, the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it (*i.e.*, experiencing the opposite outcome). A corollary to this pattern is a pattern whereby the rarer an outcome, the greater tend to be the proportions groups most susceptible to the outcome make up of both persons who experience the outcome and persons who avoid the outcome.

The patterns can be easily illustrated with normally distributed test score data. Table 1 below, which is also Table 1 of the Sessions letter, 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 columns 5 through 8) measures that might be used to appraise differences in test outcomes of AG and DG.

Column 5, which presents the ratio of AG's pass rate to DG's pass rate,⁵ 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

⁵ While I commonly refer to patterns of relative differences in this letter, the table actually presents rate ratios (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 is more common to employ the disadvantaged group's rate as the numerator for 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

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.

Table 1. Illustration of effects of lowering a test cutoff on measures of differences in test outcomes

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%

That lowering a cutoff tends to reduce relative differences in pass rates is well understood and underlies the widespread view that lowering a cutoff tends to reduce the disparate impact of tests on which some groups outperform others.

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.

Columns 7 and 8 show the proportions DG makes up of persons who pass and fail the test at each cutoff in a situation where DG makes up 50 percent of persons taking the test. Column 7 shows that lowering the cutoff increases the proportion DG makes up of persons who pass from 44 percent to 48 percent (hence, *reducing* all measures of difference between the proportions DG makes up of persons who took the test and persons who passed the test). Column 8 shows that lowering the cutoff increases the proportion DG makes up persons who fail the test from 65 percent to 72 percent (hence, *increasing* all measures of difference between the proportions DG makes up of persons who took the test and persons who failed the test).

The patterns reflected in Table 1 are not peculiar to test score data or the numbers I used to illustrate them. Rather, the patterns can be found in virtually any setting where two groups have different, more or less normal, distributions of factors associated with experiencing some outcome. Income and credit score data, for example, show how lowering an income or credit score requirement, while tending to reduce relative racial differences in meeting the requirement, will tend to increase relative racial differences in failing to meet the requirement. See Tables 2 and 3 of the Sessions letter. The information in the tables necessarily also means that lowering the requirements increases the proportions African Americans make up of persons who meet the

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 rate 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 frequency of an outcome changes, the two relative differences tend to change in opposite directions.

requirement and persons who fail to meet the requirement. Many other examples may be found in the longer references listed at the top of page 3, the scores of web pages on jpscanlan.com devoted to measurement issues, and the university methods workshops and conference presentations listed under the [Conference Presentations](#) subpage of the Publications page of jpscanlan.com.

The patterns are also evident in many types of data on school discipline outcomes, including data in DOE publications. Tables 2 through 5 below are based on data from a March 2014 DOE publication titled “[Data Snapshot: School Discipline](#).” The document provided information on the proportions demographic groups made up of K-12 and preschool students suspended one time and suspended multiple times. From the information provided in the report, one can then determine the proportions the groups made up of persons suspended (a) one or more times and (b) more than one time. Tables 2 and 3 present that information for black and male K-12 students and Tables 4 and 5 present the information for black and male preschool students.⁶

The tables illustrate the effects of relaxing standards in a way that would cause all students to receive a reprimand rather than what would otherwise be their first suspension. Such a modification would cause the proportion the indicated groups makes up of students with one or more suspensions to change from that in the first row to that in the second row. Thus, for example, as shown in Table 2, relaxing the standard in the manner indicated would cause the proportion African American students make up of K-12 students suspended one or more times to increase from 37 percent to 42 percent.

Table 2. Illustration of effect of giving all persons a reprimand instead of their first suspension on proportion black students make up of K-12 students suspended one or more times

Outcome	Black Proportion of K-12 Students Experiencing the Outcome
One or more suspensions	37%
Two or more suspensions	42%

Tables 3 shows a like pattern for male K-12 students, and Tables 4 and 5 shows like patterns for black and male preschool students.

⁶ Demographic differences in rates of experiencing things like single suspensions cannot be effectively analyzed, just as differences in rates of receiving grades of C or experiencing fair health cannot be effectively analyzed. See the [Intermediate Outcomes](#) subpage of the Scanlan’s Rule page of jpscanlan.com. It is possible that DOE has come to appreciate aspects of this issue. In DOE’s 2016 publication on school discipline titled “[2013-2014 Civil Rights Data Collection – A First Look](#),” the agency no longer presented data on single suspensions but included information on single suspensions within the category of “one or more suspensions.”

Table 3. Illustration of effect of giving all persons a reprimand instead of their first suspension on proportion male students make up of K-12 students suspended one or more times

Outcome	Male Proportion of K-12 Students Experiencing the Outcome
One or more suspensions	70%
Two or more suspensions	72%

Table 4. Illustration of effect of giving all persons a reprimand instead of their first suspension on proportion black preschool students make up of preschool students suspended one or more times

Outcome	Black Proportion of Preschool Students Experiencing the Outcome
One or more suspensions	44%
Two or more suspensions	48%

Table 5. Illustration of effect of giving all persons a reprimand instead of their first suspension on proportion male preschool students make up of preschool students suspended one or more times

Outcome	Black Proportion of Preschool Students Experiencing the Outcome
One or more suspensions	80%
Two or more suspensions	82%

If standards were further relaxed such that all persons were given reprimands for what would otherwise be their first two suspensions, the figures for the proportion black and male students make up of persons experiencing one or more suspensions would almost certainly rise still further. Rarely will one fail to observe such a pattern in circumstances where there are large numbers of observations.

In the school discipline context, in point of fact, one observes that all across the country recent reductions in discipline rates have been accompanied by increased relative racial/ethnic differences in discipline rates. See the following web pages discussing such patterns with respect to the jurisdictions indicated in the page titles: [California Disparities](#), [Colorado Disparities](#), [Connecticut Disparities](#), [Florida Disparities](#), [Maryland Disparities](#), [Minnesota Disparities](#), [Oregon Disparities](#), [Rhode Island Disparities](#), [Utah Disparities](#), [Beaverton, OR Disparities](#), [Denver Disparities](#), [Henrico County, VA Disparities](#), [Los Angeles SWPBS](#), [Minneapolis Disparities](#), [Montgomery County, MD Disparities](#), [Portland, OR Disparities](#), [St. Paul Disparities](#), [South Bend Disparities](#).⁷ These patterns are occurring notwithstanding that

⁷ These situations usually caught my attention as a result of press reportage of the fact that discipline rates had generally declined but racial disparities had increased, often while reflecting the mistaken belief that the general declines in discipline rates should have resulted in reductions in the racial disparity. Reportage that general declines in discipline rates were accompanied by decreased racial differences in discipline generally involves situations where the observers are measuring discipline disparities in terms of absolute differences between rates.

school districts may well be doing many things beyond relaxing standards in attempting to reduce racial/ethnic differences in discipline rates.

See also (a) the [DOE Equity Report](#) subpage of the [Discipline Disparities](#) page of [jpscanlan.com](#) (regarding data in a November 2012 DOE Office of Civil Rights document titled "[Helping to Ensure Equal Access to Education: Report to the President and Secretary](#)") showing that, contrary to the agency's attribution of large relative differences in adverse discipline outcomes to zero tolerance policies, relative racial differences in expulsions are smaller in districts with zero tolerance policies than in districts without such policies) and (b) Table 8 of "Race and Mortality Revisited" (showing that relative differences in multiple suspensions are larger, though relative differences in avoiding multiple suspensions are smaller, in the setting where multiple suspensions are less common (preschool) than in the setting where multiple suspensions are more common (K-12)).

These patterns, of course, will not be observed in every case, since other factors will be at work. But that does not alter the fact that general reductions in discipline rates will tend to affect measures of demographic difference in ways that are the exact opposite of what the government has been leading school administrators and others to believe. Further, the effects of the misunderstanding promoted by the government are substantial, as teachers and administrators must struggle to explain to supervisors, oversight authorities, and the public (and, in the case of agreements with the DOE, to the DOE itself) why relaxing of standards are accompanied by effects on measures of disparity in adverse discipline outcome that are the opposite of what DOE and other government agencies have led them to expect.

In these circumstances, the obligation of the agencies to correct the misunderstandings it has promoted, and to do so as soon as possible, should be evident.⁸

Sincerely,

/s/ James P. Scanlan

James P. Scanlan

⁸ One closely related matter that also requires early attention from DOE involves the agency's perceptions about the implications of the fact that students with disabilities make up a high proportion of persons subject to physical restraints. See the [Restraint Disparities](#) subpage of the [Discipline Disparities](#) page of [jpscanlan.com](#) regarding the agency's singling out of states based on the proportion students with disabilities make up of students physically restrained where the states the agency singles out favorably are those least likely to adhere to DOE guidance to employ physical restraints as a last resort, while the states the agency singles out unfavorably are those most likely to adhere to DOE guidance on the matter.