Response to December Briefing before the U.S. Commission on Civil Rights
The School-to-Prison Pipeline:
The Intersections of Students of Color with Disabilities
January 16, 2018
Additional written testimony of Daniel J. Losen

Response to Testimony of Co-Panelists and Commissioner Questions: I wish to thank the U.S. Commission on Civil Rights again for the opportunity to testify at the December 8th 2017 briefing. I submit this additional written testimony in response to questions from the commissioners and statements and research presented by other panelists.

I. Introduction: Large racial disparities in schools’ discipline of students with disabilities and the impact on instruction demonstrates the need for both the DOJ/OCR guidance on discipline and the 2016 IDEA regulations: I provide the following graph to ensure that the context of the discussion of the briefing reflects the magnitude of the disparate impact of the status quo.

**Figure 1: Disciplinary Removals For Students With Disabilities (per 100 enrolled) by Race in 2014-15**

<table>
<thead>
<tr>
<th>Removals Per 100 14-15</th>
<th>0</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>64.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>22.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: U. S. Department of Education

Figure 1 represents a racial gap of 41.5 more disciplinary removals for Black students with disabilities than for White students with disabilities (per 100 enrolled) as provided by the U.S. Department of Education for the most recent year available. The Center for Civil Rights Remedies at UCLA’s Civil Rights Project converted additional information on the removals depicted in figure 1 into an estimate of the days of missed instruction, and found that nationally, Black students with disabilities lost 119 days of instruction (per 100 enrolled) which is 76 more days than the 43 estimated to have been lost by their White counter-parts. In my written and oral testimony I explained how we derived those estimates. The national racial gap is large whether described in terms of Blacks receiving far more removals, or
resulting larger difference in lost instruction. However, there are numerous states, thousands of districts, and even more schools with far larger racial gaps.

I encourage the U.S. Commission on Civil Rights to officially acknowledge that these national data disparities support the conclusion that some schools and districts are likely removing students of color with disabilities at unjustifiably high rates.

I further urge the Commission’s report that follows to highlight the impact on instruction and recommend that federal and state agencies continue to support ongoing discipline reform efforts across the nation. The fact that some reform efforts are poorly implemented does not change the fact that in far too many schools children of color have been denied their right to an equitable educational opportunity.

The following comments are my responses to those panelists who presented research findings at the briefing as part of their arguments that either there was no disciplinary discrimination against students with disabilities and that the IDEA regulations regarding racial disproportionality in special education should be rescinded (Dr. Paul Morgan) or that the federal joint OCR/DOJ guidance that was intended to address these disparities should be eliminated for a wide range of reasons presented in the testimony of James Scanlan and Max Eden. I hope the oral and written testimony I provided on December 8th, along with my responses to information provided at the briefing and to questions asked by members of the Commission, persuades all the members of the Commission of the need to reject the status quo of huge racial disparities in lost instructional time due to discipline.

II. Areas of Agreement with Paul Morgan, James Scanlan, and Max Eden: My response to the briefing begins by providing details about numerous areas of agreement and distinctions that may have been missed or glossed over in the lively exchanges during the briefing. In the second part of this response I illustrate a number of serious flaws in the research that were presented by my co-panelists. In both sections I point out important oversights made by my co-panelists when presenting their findings or responding to comments. In the interests of keeping these responses focused, I have included technical responses and detailed examples in an appendix for those readers interested in a more comprehensive review.

a. Dr. Morgan found that Black students with disabilities had been subjected to discriminatory discipline based on race: Dr. Paul Morgan’s research findings, that Black students with IEPs appear to have been subjected to discriminatory discipline based on their race, unequivocally supports leaving the joint DOJ/OCR 2014 discipline Title VI guidance in place. His testimony also supports implementing the 2016 IDEA regulations without delay because the regulations enable districts with large disparities in discipline to address possibly racially discriminatory practices at the school and district level even if that is suspected to be a root cause, and even if the cause, once determined, is not disability based discrimination.

Among the most important aspects of the 2016 IDEA regulations is that they clarify that the law requires an inquiry into discipline disparities once they meet a threshold established by the state. Specifically, if a state finds the level of racial disproportionality in discipline meets the threshold the state created, then the law requires the district to use 15% of their Part B funds to address the root cause, or causes, with comprehensive coordinated intervening services. The actions could include training teachers to improve their student engagement,
or address a between school problem to ensure that schools with higher percentages of Black students had sufficient behavioral supports for students with IEPs. There is no directive to conduct an inquiry like Morgan’s, into whether there was proof of racially different treatment by race, or disability, because the IDEA requirement to address the root cause is a condition of receiving IDEA funds, and not based on a legal determination regarding discrimination. However, if there was evidence that suggested unintended racial discrimination was contributing to the disparities, the district could use the funds to help curb the influence of racial bias in disciplinary decisions, including for training teachers and staff to become more culturally aware and competent in working with diverse learners.

Therefore, it is important to note that Dr. Morgan testified that after controlling for poverty, prior behavior, and school-effects, his study found evidence that Black students were discriminated against in discipline. Equally important is that Dr. Morgan stated that he also found that Blacks were more likely to wind up in the juvenile justice system as a result. I call the commissioners attention to these two areas of important agreement between Dr. Morgan and myself. Where I do critique Morgan’s research, it is for employing methods that tend to under-state the degree of race-based differences. Arguably, Morgan’s research is an effort to detect race or disability bias by eliminating the amount of disparity that can be explained by another factor, such as prior behavior or lower achievement. However, several of the variables, such as prior behavior, could have been equally affected by the bias that the research is meant to detect. It is notable that despite this flaw, his findings were still consistent with the growing consensus that racial discrimination in discipline is contributing significantly to the observed disparities.

**Many studies show evidence of discrimination even after controlling for prior behavior:** I wish to call attention to the fact that the findings of Dr. Morgan’s contradict Max Eden’s expressed opinion, that there is no evidence of discrimination when similar students are compared and controls for prior behavior are added.

However, I was surprised that Mr. Eden said he agreed that I had cause to question the underlying assumption, that if one is trying to measure current racial bias one cannot assume that racial bias did not impact the reports of prior behavior that are used as the control. If the prior reports are equally biased, the research, by design is controlling for the factor it is trying to measure. One would therefore expect that adding in the control of prior suspensions would wipe away nearly all the disparities that seemed to be attributable to only racial differences. Even if only some bias influenced the prior behavioral records or evaluations, controlling for prior behaviors can cancel out a great deal of the racial disparities and skew the findings away from evidence of racial discrimination.

A more robust study, with similar findings to those of Dr. Morgan’s, comes from researchers at Tulane (and contributing authors from University of Arkansas, Rand and Brookings). The Louisiana study does attempt to account for differences in prior behavior and several other variables. Their study is noteworthy in part because they looked at 14 years worth of data from the state of Louisiana from grades K-12. Their analysis is robust, in part, because they had over 10 million student observations and had nearly all the student records from every school and district. Unlike Morgan’s sampled data, the Louisiana data set enabled the researchers to look at both school level and district level factors. Perhaps most important is that they found *both* “within-school” and “between-school” evidence of
discrimination. Despite the opinions expressed by Max Eden, consistent with Dr. Morgan’s findings, these researchers found that within-school factors explained the greatest share of the discipline disparity.

The Tulane study also examined differences in the severity of punishment for fighting. The authors expected that adults would most likely mete out fair responses in fights between Black and White students, yet they did still detect racial differences, albeit small ones, in how the adults responded to fights, controlling for prior behavior. Although beyond the scope of their inquiry, the Louisiana study authors did believe that they would be more likely to find racially discriminatory treatment in the minor and more subjective categories.

In contrast to the comprehensive nature of the Louisiana study, Dr. Morgan’s methods eliminated all discrimination that may have been caused by policies and practices. These are called “between-school” factors, and this method is also known as controlling for “school-effects.” It is critically important to note that Chairperson Lhamon and Dr. Morgan agreed that his analysis was designed to compare similarly situated students and control for as many other contributing variables as possible. This limitation applies to both Dr. Morgan’s disciplinary data analysis and his analysis of differences in identification for special education. Therefore, I caution the Commission from giving too much weight to Dr. Morgan’s findings with regard to identification as well as discipline because his research was designed to only detect different treatment of similarly situated students.

Further, in my oral testimony I emphasized that it is district and school level differences that are the primary subject of the IDEA’s regulations on disproportionality in special education. Likewise, policies and practices that have a racially disparate impact on students at the school and district level are also a critically important part of the federal OCR/DOJ guidance. Yet, none of Dr. Morgan’s numerous studies entails an analysis of district level disparities. This is relevant because where Dr. Morgan found no evidence of discrimination by disability status in the area of discipline or by race in the area of special education identification, his null findings are partly due to the fact that national samples do not come close to the level of disparities experienced by children at the school and district levels. I have been on panels with Dr. Morgan before and we have agreed that his research conclusions cannot be generalized to all districts, especially not those whose racial disparities differ from Dr. Morgan’s sample by orders of magnitude.

Unfortunately, there are numerous districts with very large racial differences in discipline among students with disabilities at the district level. These are documented in our report entitled “Are We closing the School Discipline Gap” which highlights Memphis Tennessee to illustrate just how large a difference there can be. Specifically, on page 8 of “Are We Closing the School Discipline Gap” we point out that 46% of Black secondary students with disabilities were suspended out-of-school at least once versus just 16% of White students.

---

1 The full technical brief for the report is available at

2 I’d like to submit the Tulane Center’s report and technical brief for the record as it is among the most recent and most robust available. Barrett, N., McEachin, A., Mills, J.N. and Valant, J., 2017. Disparities in Student Discipline by Race and Family Income. Education Research Alliance for New Orleans.

3 Ibid., P. 30. In the conclusion of their (brief or technical report) they opine that their findings likely indicate greater racial bias in how punishments are meted out for less serious offenses.

4 The report has since been peer reviewed by the American Education Research Association and awarded the “Outstanding” policy report of 2016.
That means that Memphis’ racial gap of 30 percentage points is nearly double the national racial gap between these two groups.  

My response raises this issue and will attach the report because at least one commissioner asked for additional empirical data that conflicted with Dr. Morgan’s findings. We can safely assume that the level of disparity found in Dr. Morgan’s sample is closer to the national average than it is to the levels of disparity observed in Memphis. I am certain that after looking at the district level disparities, Dr. Morgan would agree that the national sample he examined and presented to the Commission had far lower racial discipline disparities than all of the districts we deemed to have large disparities in our report. I present Figure 2 to illustrate that there are districts with secondary suspension rates that are so high, and with racial disparities so large, that they shock the conscience. The rates below are out-of-school suspension rates for Black and White secondary students with disabilities from five districts we deemed to be among the high suspending districts in “Are We Closing the School Discipline Gap?”

**Figure 2: District Disparities From Those Featured in Are We Closing the School Discipline Gap?**

![Figure 2: District Disparities From Those Featured in Are We Closing the School Discipline Gap?](image)

It is well established that disparities by themselves are insufficient to conclude that there is unlawful discrimination. But if the federal government rescinds either the DOJ/OCR discipline guidance or the IDEA 2016 regulations, such actions will increase the likelihood that even the most extreme disparities at the district and school levels are ignored. Even though all agree that the research provided by Dr. Morgan was not created to explore district level racial disparities in discipline of this magnitude, one of my greatest concerns about the discourse at the Commission’s briefing is that some might assume that Dr. Morgan’s limited research findings of sampled national data accurately reflect the magnitude of the district and school level disparities and the need to address them (in discipline and identification). Specifically, such national-level findings should never justify

---

diminishing district-level oversight. I hope the Commission agrees that the need for district level interventions is only fairly captured by a district level review.  

b. Areas of Agreement with Max Eden Who Documented Successful Discipline Reforms in New York City: Ironically, a point of agreement I have with Mr. Eden comes from the findings of his own study of New York City reforms. I agree with Eden’s conclusions that reforms that Mayor Bloomberg pursued lowered suspensions and did not coincide with a district-wide degradation of school climate. I also agree that mandating schools to not suspend students out of school for minor behaviors for their first offense, as Mayor Bloomberg did, is a reasonable alternative where suspensions are excessive and disparate. The commissioners should take note of two other areas of agreement that I have with Max Eden’s report. The report described two other reforms contained in guidance from Mayor Bloomberg that coincided with a decline in suspensions with no deterioration of the school climate. Specifically, although Eden’s oral response only mentioned the one reform, his report also describes other Bloomberg reforms, such as encouraging schools to shorten the days of suspension as follows:

The second was that for students in kindergarten through third grade, the maximum suspension was reduced from 10 to five days for mid-level offenses categorized as “disruptive behavior,” such as shoving a fellow student, using a racial slur, or engaging in inappropriate physical contact. The guidance also informed teachers that a “restorative approach can be used as both a prevention and intervention measure.”

Therefore, it is reasonable to argue that even Eden’s analyses supports the concept that large urban districts could be inspired by the joint OCR/DOJ guidance to replicate the three Bloomberg reforms — eliminate suspensions for first time minor offenses, encourage teachers to pursue restorative justice practices, and mandate shorter suspensions, — lowering the disparate impact of lost instruction.

In New York City, reform efforts under both Mayor Bloomberg, followed by Mayor DiBlasio, lowered the number of suspensions by approximately 16,000. Max Eden’s report on these efforts distinguishes Mayor Bloomberg’s efforts as successful, while harshly criticizing those under Mayor DiBlasio. My serious disagreements with the report’s findings about the school climate declines under Mayor DiBlasio are addressed in Part II of my response.

Citywide New York Data on Teacher’s Sense of Order and Discipline Are Important to Review: I also agree that school climate surveys are important to review when evaluating the efficacy of discipline reform. I think it is particularly important to reflect on the substance of teachers’ responses. In particular the school survey prompt that NYC asks teachers to respond to each year is one that should be paid attention to. The prompt reads: My school maintains good order and discipline.

Teachers are asked to strongly agree; agree; disagree; or strongly disagree. The four other survey responses analyzed in Eden’s New York City report are structured in a similar way

---

6 Therefore I am also submitting with my response a link to our report, Are We Closing the School Discipline Gap? as well to the accompanying spreadsheet containing the analysis of the race and disability differences in suspension rates for every district in the nation.

and they are from students. One asks about fighting; another about gangs; a third about drug use, and the final one asks whether students show mutual respect for each other.

Most important is that Eden’s conclusions in his testimony, that his New York City analysis suggests chaos resulted there, are not supported by the details of his own report because, as he states in the text, “...in both Period 1 and Period 2, the distribution of differences between schools with neutral suspension rates and those with declining suspension rates was similar for all questions.” In other words, Mr. Eden confirmed in his testimony that his analysis found no correlation at the school level between reducing suspensions and a diminishing school climate for any of the five indicators he studied.

Ideally, one would expect a report on climate to at least track the percentage of positive responses and observe trends over time. Unfortunately, Mr. Eden’s NYC report provides neither citywide nor even individual school level responses on this survey item, nor are the levels provided for the four other survey responses that his report analyzes. Instead his report looked only at the changes in the value of the responses. This is arguably a fatal flaw.

Figure 3 shows the publicly available city-wide summaries of the teachers’ responses to the prompt: My school maintains good order and discipline. The graph represents seven annual surveys between 2010-11 and 2016-17 and the data are from publicly available summaries produced by the New York City Department of Education. The most striking aspect of these stable and arguably positive citywide results is that Max Eden analyzed the same raw data and testified before the Commission that he saw evidence of “catastrophic” failure.

Figure 3: Citywide Responses of NYC Teachers Shows Calm More Than Chaos

Source: NYC School Survey Results

---


9 The NYCSDD presentation of the data was excerpted from two presentations: The data from 13-14 to 15-16 comes from the school district’s website here: http://schools.nyc.gov/Accountability/tools/survey/2017Survey.htm and the report link is here for the most recent review including 2016-17. http://schools.nyc.gov/NR/rdonlyres/53642445-2C99-4AC5-A0F9-E34ACFC25B88/0/2017CitywideAnalysisofSurveyResults.pdf

The data in slide 11 are presented as representing the response strongly agree/agree “Order and discipline are maintained at my school”
The columns in figure 3 show the actual response levels to the question posed to the teachers as reported by the NYC Department of Education. Most researchers would acknowledge these values since they came from the identical data source. Eden's testimony, that "we should listen to teachers," compelled me to compare his findings with the actual New York City teacher survey responses regarding their impression of order and discipline in their school. The Commission should note that the citywide results published by the City of New York directly conflict with Eden’s stated opinion about his findings.

Although he does not comment on the actual levels for this indicator (or any of the others) in his report it does include a table (Figure 6) where one can see that throughout the entire period, at least two thirds of NYC’s schools had less than 30% of teachers giving a negative response to the prompt about maintaining order and discipline.

It should be noted that he 2017 NYC school climate survey also polled parents to learn their top requests for school improvements. There were 9 total top requests. The top 3 requests were for stronger enrichment programs (24%), smaller class sizes (21%), and more hands on learning (15%). The request for a safer school environment (5%) tied with stronger school leadership (5%) and represented the lowest school improvement priorities for parents.  

In the appendix I break down the unique methods for the NYC findings and how they could lead to a very different set of conclusions. Simply put, Eden’s unique methods of his own design systematically discounted survey results that showed consistently strong school climate from one year to the next. By prioritizing the number of schools that showed a change from one year to the next, but only counting changes that reached a certain magnitude, Eden’s research design failed to credit the schools whose highly rated climates stayed highly rated throughout the period. Similarly, his methods failed to capture those schools that showed small levels of improvement.

Ultimately, the fact that the underlying and contradictory values were over-looked, and that the schools that maintained strong school climates were devalued by the study’s emphasis on change, casts serious doubt on the validity of Eden’s conclusions about New York City’s school climate and undermines his claims about observing "chaos" in other cities.

---

[Hyperlinks to sources provided at the bottom of the page.]
There is strong evidence that school and district policies and practices cause a large portion of the racial differences in discipline: Finally, I agree with Max Eden’s statement to the commissioners when summarizing other studies: “It's notable though that all these studies, ... has found a consistent pattern that schools with higher percentages of African American students have higher suspension rates. This could be perceived as troubling and could be interpreted as systemic discrimination or systemic bias. " Along these lines, it is noteworthy is that both Eden and I acknowledge that the findings from the study of Louisiana are important. Eden and I might further agree as to the relevance of racial differences driven by both within-school and between-school factors if not for his erroneous legal understanding revealed in his oral testimony, “Schools cannot discriminate, only people can discriminate.” The law directly contradicts this opinion that only the different treatment of students by individual teachers or principals can be considered discriminatory.

For example, it is well established that the equal protection clause of the United States Constitution applies to a wide range of state action, including school and district policies and practices. If a school district intentionally set harsher discipline policies for schools serving higher percentages of Black students it would likely violate the Constitution as well as Title VI protections pursuant to the Civil Rights Act of 1964. If such school level policy differences developed without conscious intent, but had a racially disparate and harmful impact, the guidance makes clear what the regulations have always said, that such policies may still implicate civil rights protections if the differences can not be justified.

c. Areas of Agreement with James Scanlan's testimony: James Scanlan offered several technical observations with which I agree. Moreover, if we agreed about what the law requires, our areas of disagreement would be further diminished.

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 racial differences narrow and suspension rates decline. This means that interventions that have helped to lower Black suspension rates more than they helped to lower White suspension rates can still look like they made racial disproportionality expand, but only if one restricts their review to relative risk ratios. The reason for this counter-intuitive tendency is that when two values are both declining, the relative ratio of the amount of decline must exceed the starting relative risk ratio. If the ratio of decline is smaller, the new risk ratio will always be larger than the risk ratio one started with.

Mr. Scanlan overlooked the obvious solution, which is to pick a tool to measure racial disparities that doesn’t have the problematic tendency that Scanlan highlighted for the Commission. For example, I found that flipping the same data over to examine the risk ratio of the non-suspended solves the original problem. I have observed that when suspension rates decline and the racial gap narrows, if the absolute value of the decline is greatest for the highest suspended group, the non-suspension ratio invariably improves (gets smaller not wider) and this is consistent with a narrowing of the racial gap in absolute terms. As part of CCRR’s ongoing work with educators my organization has encouraged states and districts to look at the change in the racial gap (the absolute difference) rather than rely on relative racial disparity metrics to assess racial progress in discipline over time.
There is nothing new about looking at racial differences in absolute rather than relative terms. Educators are familiar with efforts to close the achievement gap, and the graduation rate gap. The discipline gap is no different.

Unfortunately, Mr. Scanlan misunderstands what anti-discrimination investigations can and should consider as potential evidence of discrimination. Ultimately, Mr. Scanlan’s testimony is not relevant to the OCR/DOJ guidance because neither the discipline guidance nor anti-discrimination law requires the use of relative risk ratios. However, his observations are relevant to the 2016 IDEA regulations, as risk ratios must be used pursuant the new regulations. However, the regulations do not (and should not) restrict districts from considering the non-suspension risk ratio, which can also be described as the inverted suspension risk ratio, as long as reasonable thresholds are established and the state provides sound justification for where they are set. In appendix A of this testimony I present a table showing how the non-suspension risk ratio, as well as the suspension risk gap each avoid the problematic tendency that Scanlan has observed with the relative risk ratio.

**Part II Flaws and other concerns with the Testimony of James Scanlan, Max Eden, and Paul Morgan**

**a. Flaws in the testimony of James Scanlan:** By insisting, despite evidence to the contrary, that there is only one legitimate way to measure racial disparity in the education context, Mr. Scanlan’s testimony was built on an illogical assumption. The fact remains that there are numerous acceptable ways to measure racial disparity in discipline for the purpose of anti-discrimination law enforcement. Although federal and state agencies should be expected and able to justify why a certain measure is being used, the method of measuring racial disparities should reflect the given context and serve the interests of fundamental fairness and preventing racial injustice.

James Scanlan argues that the joint DOJ/OCR guidance is wrong to encourage reductions in discipline because of the observed tendencies of risk ratios to give conflicting signals in some situations. His presentation ignores a central tenet of disparate impact theory, that one of the protected classes must be harmed by the policy or practice more than others. Logic requires that a less discriminatory alternative would lower the harm experienced by the group with the higher suspension rate. All agree that a ratio can increase or decrease in size without indicating whether the harm to the higher suspended group has diminished.

The Commissioners should note that the U.S. Department of Justice and the Institute for Education Sciences (IES) recognize that looking at the racial gap in suspension rates is a viable method for measuring and monitoring racial disparities in school discipline, and one that doesn’t share the problematic tendencies of the relative risk ratio. In the appendix I provide an excerpt from a technical assistance brief designed to help districts review racial disparities in discipline data. It was produced by WestEd for the North East Regional Education Lab, reviewed and approved for publication by the current administration and published last spring. The guidance demonstrates that the short-comings of suspension risk ratios for measuring racial progress in school discipline over time, and that there are

---

available alternative measures that can be used, are both well-established and recognized by the IES.

b. Misinterpretations and Polarizing Rhetoric in Max Eden’s Testimony Should Not Go Unnoticed: I encourage the Commissioners and the public to consider Max Eden’s polarizing National Review polemic subtitled, Are America’s Teachers a Bunch of Racists? Democrats Seem to Think So. In this piece, as in his oral testimony, Eden suggests that all those who call attention to racial injustice in discipline, (including former Secretary of Education Duncan) are “slandering” teachers.12 Max Eden’s several published opinions attempt to paint discipline reformers and the DOJ/OCR guidance as “anti-teacher.” The available facts tell a different story. For example, last year the American Federation of Teachers dedicated an entire issue of their monthly magazine to supporting discipline reform, and both the NEA and the AFT have issued statements acknowledging the need to address racial disparities by changing policies and practices.

I find it ironic that his testimony suggesting that the guidance was motivated by a belief that all teachers intend to discriminate against children because the guidance explains the circumstances under which a disparate impact of a policy or practice might be regarded as unlawful. Dis disparate impact prompts educators to rethink otherwise race-neutral policies or practices if they are found to have an unintended harmful disparate impact. The guidance provides numerous examples to make this clear. Not one of the examples in section two of the discipline guidance concerns situations where the disparities are directly caused by teacher or principal bias.13 However, in his effort to persuade the Commission that research has found little evidence of discrimination he referenced studies that found that differences between schools, also known as “school effects” account for the observed racial disparities. The disparate impact aspect of the guidance explains that racial disparities alone are not proof of discrimination, but that some school policies or practices that cause such disparities might be unlawful if they lack adequate justification. The guidance encourages schools to review their data, and to replace those policies and practices that are causing the disparities that they cannot justify with educationally justifiable ones.

Most would agree that schools and districts should be discouraged from implementing policies and practices that have negative consequences if they cannot be justified on educational grounds. This represents a common sense principle about school policy, and is one that could help all students generally. The wisdom of the approach doesn’t change simply because it is only enforceable under civil rights law when unjustifiably bad policies impact protected subgroups of students more than others. Unfortunately, removing the guidance would only serve to suggest that punitive policies that reduce instructional time need no justification, even if they harm groups of children that have historically been discriminated against.


13 It is conceivable that implicit or explicit bias may have influenced the decision to adopt a particularly strict policy, or to maintain a policy that was counter productive, such as suspension for being tardy. Perhaps the policymakers did not realize that it was hurting one group more than others and not achieving the intended goal, yet when those problems were brought to their attention they maintained the policy anyway. Absent clear proof of racist intent, the guidance makes it clear to school leaders that an unsound policy might still be unlawful because of its harmful disparate impact.
A Great Deal of Important Information is Over-stated, Misstated, or Over-looked in Eden’s Testimony: I would also like to call the commission’s attention to Mr. Eden’s assertion during his oral testimony that the effort to address the school to prison pipeline “is creating a school climate catastrophe and putting at risk students at greater risk.” Neither this statement, nor his testimony that all the studies that have said suspensions predict harm assumed causation from mere correlation, is supported in the evidence.

In my written testimony I provided a number of citations to robust research papers that were not merely correlational, but did control for prior behavior, poverty, district effects and many other variables to predict lower graduation rates, lower achievement and dramatically increased for involvement in the juvenile justice system, as much as threefold greater. I’d encourage the Commission to recognize that the non-partisan Council of State Governments Justice Center working with Texas data produced the most rigorous study of the harms from school discipline and contributors to racial disparity in the field (Fabelo, 2011). The researchers had data from every middle school student in the state and tracked 3 middle school cohorts for over 6 years.

Their findings inspired the creation of the federal government’s Rethink School Discipline Initiative in 2014, and contributed to the perception that issuing the Joint OCR/DOJ guidance letter as part of a package of technical assistance would help districts and students.

Additionally, the “Rethink School Discipline” initiative resulted in a non-partisan Council of State Governments consensus report, based on research and the input of teachers unions, school administrators, law enforcement, juvenile justices and researchers. The Consensus report was published in 2016. The CSG Consensus School Discipline Report acknowledges that racial disparities are often indicative of a problem that schools can remedy. I’d also point out that the Texas-based Council of State Governments Justice Center, whose central mission is to provide practical, nonpartisan, research-driven strategies and tools to increase public safety and strengthen communities, recently published Realizing the Full Vision of School Discipline Reform: A Framework for Statewide Change (September 2017) which includes a section promoting efforts to reduce racial disparities in discipline.

Fortunately, several states have followed the kind of common sense discipline reform recommended by the consensus report, the guidance, and the wealth of research. For example, Republican Governor of Texas, Greg Abbot inspired by what he considered to be successful initiatives in Houston and Dallas recently signed state legislation that banned suspensions for grades K-2 for minor non-violent behaviors. Republican legislators in Ohio just introduced a similar piece of legislation. I mention the efforts lead by Republicans because eliminating unnecessary and unjustifiable racial disparities and ensuring that disciplinary exclusion is a measure of last resort is grounded in good common sense, not

---

14 Eden’s inference of direct causation can be found at the 1:50 mark in his oral testimony here: https://www.youtube.com/watch?v=yv_GgG-ig8E&t=63s
politics. The reason that discipline reform has resisted becoming political, at least until now, is that there are far too many good reasons to try and reduce suspensions.

As a factual matter, the guidance is purely explanatory. However, Mr. Eden’s testimony claimed that the guidance dictates specific reforms yet offered no specific examples from the guidance. Further, his testimony suggested that the guidance has resulted in the federal government accusing school districts of racial discrimination strictly on the basis of data disparities, yet it provided no documentation to support the assertion. For these reasons I conclude that Mr. Eden’s testimony about the requirements of the guidance misrepresents their actual contents.

His testimony also raises serious doubts in my mind regarding the credibility of his summary of research findings. One important example is the frequent claim that there are no robust studies suggesting harm from discipline. That is clearly debunked by the aforementioned CSG study that tracked every middle school student in the state of Texas for over 6 years and was hailed for its rigor by Republican state legislators. I contrast this claim with his related citation in his oral testimony of the findings from a University of Arkansas study that he has repeatedly called the “most robust” study, and one that his testimony suggests provides causal proof that getting suspended improved test scores and is therefore helpful, not harmful.

I have published a review of the research in question with another researcher, Dr. Brea Perry and we each roundly criticized the purported “robust” University of Arkansas working paper Eden referenced because its methods were flawed, its findings were of dubious validity, and because they failed to meet research standards for “robustness.”

Our critique further warned the authors that in several places their summary writing could lead to the kind of misinterpretation that Max Eden has since engaged in. The authors’ response stated that they thought their findings were clear. They have subsequently sought to correct the highly publicized misstatements about their findings. First, they did tweet a response directly to Max Eden’s colleague at the Manhattan Institute, Jason Riley, insisting that Mr. Riley’s Wall Street Journal’s OpEd overstated their research findings. More recently, the authors tweeted directly to Max Eden, well before he testified on December 8th, telling Mr. Eden that their study, which only examined the lagged impact of suspension on test scores of suspended students a year or more after they had been suspended, did not support a conclusions that suspensions improved test scores, but merely that they were not able to discern a negative impact from being suspended a year or more later. In response to their tweet, Max Eden insisted that he had a right to interpret their study findings as he saw fit.¹⁷

Some of the core substantive criticisms that Dr. Perry and I described in our published review of their study included: the University of Arkansas researchers could have but failed to attempt to answer the most obvious question, “What was the impact of suspension on test scores the year the student was suspended?” One would think that if getting suspended caused a student to miss a week of Algebra class, it would much more likely impact their Algebra test score the same year the student got suspended, than it would impact a Geometry test the student took more than 12 months later, when the student missed no Geometry instruction. By only focusing on the lagged impact, the research design ignored the substantive achievement impact from missing instruction

¹⁷ Links or copies of this exchange on Twitter are on file with the author and can be made available upon request.
in a particular class or subject. Max Eden’s oral testimony and written references fail to mention the key limitations to the unpublished working paper.

Equally problematic was that the complex research model the researchers used in attempt to approach a causal analysis of the lagged impact on test scores, excluded all students who experienced the greatest long-term harm including: suspended students that subsequently dropped out, and those that were sent to the juvenile justice system or an alternative school. In other words, when Max Eden suggests that this working paper provides “the most robust” counter-example to studies that predict harm from discipline he fails to mention that his preferred study excluded all the students who experienced the most predicted harms.

The study also failed to control for students who repeated their grade and were being tested on the same material they had covered the prior year. Finally, with regard to how reliable the study findings were, even in terms of the lagged impact, my fellow reviewer, Dr. Brea Perry pointed out that their first posted draft reported the levels from the standard statistical robustness check, and these levels indicated that their study findings failed to meet the research robustness standards.

Max Eden concluded his oral testimony by telling the Commission that wherever he looks he has seen that discipline reform has made schools unsafe. I hope that my response to his testimony has shed some light on the numerous findings and contradictory data that Mr. Eden has overlooked in forming the opinions he has offered to the Commission.

**Conclusion: There is no justification for delaying or eliminating our children's civil rights safeguards against discriminatory removal from school.**

My response testimony has aimed to eliminate the erroneous assumptions and misunderstandings about the law that fuel the arguments against the federal government providing oversight and assistance to districts that are critical to protecting children from discriminatory policies and practices.

One topic mentioned by Panelist Dr. Monique Morris, but not explored in depth during the briefing, is that different forms of racial discrimination can exist simultaneously and can overlap and influence each other. Although research on the ways implicit bias may influence the development of harsher discipline policies and practices is beyond the scope of my response, this possible overlap should be acknowledged. Fortunately the joint OCR/DOJ guidance and the 2016 IDEA regulations prompt schools to address both individual discrimination and the discriminatory disparate impact of policies or practices. This overlapping protection is important, as is the possibility that implicit unintended racial bias on the part of educators can influence both types of discrimination.18

Most important, despite our areas of disagreement, each of the researchers who presented on Panel II acknowledged at some point during the briefing before the Commission that some degree of racial discrimination likely contributed to the observed racial disparities in discipline. The disagreement was not whether discrimination existed, but how much of the observed racial disparity was due to discriminatory different treatment, and how much was due to systemic impact of policies and practices. While the exact amount of unlawful racial

---

18 For example, the decision by a district’s board of education to allow schools serving higher percentages of Black youth to promulgate unsound discipline policies like corporal punishment, or suspensions for truancy, or to receive lower levels of support resources per pupil may reflect unconscious forms of racial bias.
discrimination was difficult to quantify, none who testified against these federal safeguards ever explained why any of the children who were discriminated against should have their protection against unfair treatment taken away.

For these reasons, and in the face of the incredibly large racial disparities in removals and lost instruction discussed at the outset, I urge the U.S. Commission on Civil Rights to state that none of the research and arguments made during the briefing provided a sufficient reason to delay or take away protections against unfair or unsound disciplinary policies or practices that threaten to harm children of color and children with disabilities.

Thank you for the opportunity to respond further to the questions and issues that arose during the December 8, 2017 briefing.

Sincerely,

Daniel J. Losen

Appendix:

a. Examples of more effective ways to measure racial disparities that do not exhibit the problematic properties described in James Scanlan's research:

In this table, I demonstrate that either the racial gap or the risk ratio of the non-suspended could more effectively register whether harm was reduced. I recommend the racial gap in suspensions be used for measuring the racial disparities in suspension because it can also be used for analyzing differences in days of missed instruction and differences in suspensions per 100 enrolled. Both of these suggested measures might help educators understand the degree and impact of the disparities with the difference in lost instruction best for demonstrating the disparate impact on learning. One could also look at the racial gap among the non-suspended, but it is identical to the suspension gap.

Table 1. Alternatives to suspension relative risk ratios

<table>
<thead>
<tr>
<th>Year A</th>
<th>Risk for Suspension</th>
<th>Suspension risk ratio</th>
<th>Risk for non-suspension</th>
<th>Non-suspension risk ratio</th>
<th>Racial Gap in Suspension Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>80%</td>
<td>80/40</td>
<td>20%</td>
<td>60/20</td>
<td>80-40</td>
</tr>
<tr>
<td>White</td>
<td>40%</td>
<td>2.0 = Blacks 2 times more likely to be suspended than Whites</td>
<td>60%</td>
<td>3.0 = Whites 3 times more likely to not be suspended than Blacks</td>
<td>40 points higher for Blacks</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year B Change</th>
<th>Decreased</th>
<th>Increased</th>
<th>Decreased</th>
<th>Decreased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>50%</td>
<td>50/20</td>
<td>50%</td>
<td>80/50</td>
</tr>
<tr>
<td>White</td>
<td>20%</td>
<td>2.2</td>
<td>80%</td>
<td>1.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year C Change</th>
<th>Decreased</th>
<th>Increased</th>
<th>Decreased</th>
<th>Decreased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>20%</td>
<td>20/5</td>
<td>80%</td>
<td>95/80</td>
</tr>
</tbody>
</table>
Table 1 above shows that two out of three measures of racial disparity decreased as the level of harm for both Blacks and Whites decreased. As demonstrated, the non-suspension risk ratio gets smaller when suspension rates decline for both groups and unlike its counterpart, the non-suspension ratio always gets smaller if the absolute decline is greater for the highest suspended group. One caveat is that if an intervention were to help both groups, but the lower-suspended group was helped much more, it is possible that the non-suspension risk ratio would get smaller even if the uneven benefit caused the absolute size of the racial gap to increase.

With regard to cautions about using risk ratios mentioned in the text, the recently published guidance states: “There are several limitations to risk ratios. First, high ratios are possible even when the underlying risk for each of the comparison subgroups is low. For example, a relative risk ratio of 2:1 could be found in a district with few suspensions or a district with many suspensions. In figure 1 the fictitious data produce the same risk ratio for four districts, even though in city A the rate of out-of-school suspensions is 80 percent for Black students and 40 percent for White students, while in city D the rate is 20 percent for Black students and 10 percent for White students. Both yield a risk ratio of 2:1, even though the absolute racial gap is much larger in city A.”

**Figure 4: Use of Relative Risk Ratio to Analyze Out-of-School Suspension Disparities Over Time Can Mask Racial Progress**


Relative risk ratios can obscure whether suspension is used frequently or as a measure of last resort. The guidance goes on to state: “Second, relative risk ratios do not indicate whether a disciplinary action is used often or rarely, which may be an issue if the goal is to discover whether large differences are driven by potentially excessive or unnecessary disciplinary actions. Third, because risk ratios depend on division, if the comparison subgroup has no suspended students, a ratio cannot be calculated. This may be the case for some elementary schools. The risk ratio should not be relied on as a primary measure for
understanding disproportionality in the district but should be used in conjunction with other metrics, such as the risk difference.” It is also worth mentioning that the recent robust study of racial disparities in suspensions by collaborators from Rand; Brookings; Tulane and the University of Arkansas analyzed racial gaps (not risk ratios) extensively throughout their technical analysis. 19

The non-suspension risk ratio is not perfect, but neither are racial gaps. The fact is that there is no definitive method established by the research community, and that is why the U.S. Department of Education has provided guidance on how to use a variety of methods. In other words there is more than one way to demonstrate that racial disparities are improving.

b. Further breakdown of concerns with Max Eden’s NYC report.

I did find several attributes of Eden's analyses that likely explain how his method describes degradation even though the findings reported by New York City describe relative stability.

Readers should note that Eden's methods are of his own creation. Questions about While the phrase “distribution of differences” appears to be similar to “differences in differences,” a statistical method designed to show causal relationships, Eden acknowledges in the report that his analysis is purely descriptive. This means that the data he described represented neither causal nor nor correlational findings.

Most important, for each indicator, Eden fails to answer the most basic question, “Did the citywide results show an actual decline?” Unfortunately, anyone who skims his detailed report might assume that that the teachers and students were asked about how much their school improved or declined. This is not the case. Survey respondents were only asked (for these indicators) about the current year, yet his analysis appears to have obscured the substance of the responses. Instead, the report provides a complex evaluation of how each set of school responses differed from responses from the prior year.

Equally important, the report created evaluation categories, and cut scores for the degree of change that “counted.” These arbitrary benchmarks were created specifically for this report and not well justified in the text. The system focused on change at the school level. However, all schools that did not change by 5 percentage points or more were labeled “similar” and thereby valued equally in the analysis. This means that all schools that teachers consistently gave 100% positive ratings were regarded as equal to those with the lowest, yet most consistent, negative ratings, so long as the responses were stable. Likewise, schools that improved or declined, but by less than 5 points were also labeled “similar.” This means that a school that made slow gradual improvement of one percentage point each year would be listed as making no change and labeled “similar” over any 2-3 year period. One can imagine that if 100% of the schools in New York improved by 4 percentage points on each of the selected indicators. Most would report such results as representing improvement but this system would register no improvement to school climate in that situation.

To illustrate the problem with this analysis using a more realistic example, imagine if 50% of the schools showed perceptions of order and discipline by teachers were between 95% and 100% favorable and these all stayed in that high range during the period in question. Then imagine that another 30% of all the schools were not in that highest range and improved their climate, but by only 2-4 percentage points. Because the combination of the high but stable schools with the improving schools that fall short of the required 5 percentage points are labeled “similar,” the analyses would show that 80% of schools stayed “similar.” And if out of the remaining 15% of schools, 5% showed improvement that was strong enough to count using the report’s scale, and 10% had lower climate ratings large enough to count, the system would show this as evidence of overall deterioration because the number of schools that got worse was twice the number that improved.

Therefore, in designing a system that disregarded schools that maintain a high score, and also positive changes of low magnitude, it was predictable that this method would overlook even strong and consistent evidence of progress and not give much credit to schools that maintained very positive climates.

The method might be fair if small changes in either direction tended to balance out. But if many more schools tended to improve a little each year, negating small changes would predictably cancel out many more signs of improvement than deterioration. One cannot tell if this was actually the case without re-running his study, but it is clear that the evaluation method was insensitive to small changes. One can see this because on nearly every indicator the category of “similar” is the largest category. As mentioned, the author provides no research basis to support the decision to not count a change of 4.9 percentage points as an improvement (or decline) but to count all changes of 5 points or more.

Having looked at the actual citywide results for teachers’ sense of order and discipline, which showed little change during the period, the likely reason for the contradictory findings is that the Report’s system magnified the relevance of large changes while ignoring positive yet high results. If we assume the official citywide survey results are accurate, then report’s method obscured the evidence of stability by focusing only on the changes at the margins.

Digging even deeper I discovered what appears to be a built in bias that is more likely to record deterioration than progress if the starting values are relatively high. A look at the starting points confirms that most schools started with survey results that were labeled good if they received less than 15% negative responses. The schools doing really well (less than 5% unfavorable) at the starting point could only register as staying the same or declining because of the requirements of the evaluation system. Readers should note that if less than 7 in 10 teachers or students didn’t rate a question positively, that school was considered to be in bad shape. The use of a 5 point and 15 point threshold for change, and lack of adjustment for school size could have further distorted the report’s findings. For example a large high school that transformed itself from 80% perceiving a gang problem to 0% gang involvement would be reported as equal to a small elementary school of 100 students that improved but by a much smaller degree of 16% negative to 0%.

Table 6 in the report is the closest it comes to providing a sense of the underlying values. On the question of Teacher’s perception of order and discipline, Table 6 shows an increase of 12 schools where more than 3 out of 10 teachers gave negative ratings of their school. As a percentage of all schools with negative rating the change went from 30.9% in 2011-12 to 32.2% in 2015-16. We are not told the magnitude of the change citywide or how much of a
decline occurred in those 12 schools. Similarly during this same period the number of schools in the top tier increased by from 436 to 438. The answer to the net citywide change in NYC’s published data is that teacher responses showed an improvement from 78% to 81%. In other words, the net declines reported by Eden must have more than made up for the improvements in the other schools that were likely too small for Eden’s system to detect as positive change.

In his report’s introduction Eden states that there was, “Less order and discipline: In 2015–16, a higher percentage of teachers—across 636 schools serving 376,716 students—reported that order and discipline were not maintained in their school, compared with two years earlier (2013–14).” What is disturbing is that the City-wide report shows a 1 percentage point gain during this period from 80% to 81%. A layperson would naturally think that both statements cannot be true, but if one counts the number of schools that changed and the other the aggregate citywide survey substantive results, it is possible.

Before expressing such strong and dramatic conclusions about the current state of school climate in NYC’s schools Eden should have reviewed the newer indicators to see if they aligned with or contradicted his conclusions. I was shocked to find that New York City reports on about 100 different indicators of school climate. Here are some other recent summary findings that Max Eden overlooked.

According to the December 2017 report 96% of parents said at their school their child is safe at school.20 79% of the students said that gang activity is rare or none. 84% of students said they felt safe in their school hallways, bathrooms, locker rooms and cafeteria, 85% said they felt safe going to and from school and 91% said they felt safe in their classes. Finally, 78% of teachers agreed that “the Chancellor is an effective manager who makes the NYC DOE run smoothly.” At the same website the district has stated that graduation rates have improved.

Finally, Eden also attributes all changes in value as caused by some distant act such as the joint DOJ/OCR guidance or Mayor Di Blasio’s due process procedures, without exploring more logical explanations. For example, a quick search revealed news reports of increasing gang activity in New York City during the period in question. Although I didn’t attempt to confirm this by finding gang statistics, it’s quite clear that Eden didn’t either. If true, one would think that and increase in gang activity would be a more logical reason that students reported an increase in gang activity and fighting at their school than a change in the suspension due process procedures. So if schools did experience an increase in violence, the methods in the report provide no way to distinguish the impact from the policy in question, from a genuine rise in gangs, caused lower safety ratings.

_____________________________________

20 See tool and click additional survey questions at https://tools.nycenet.edu/guide/2017/#dbn=City&report_type=ALL