



UNIVERSITY OF
ARKANSAS

College of Education & Health Professions
Education Reform

**Understanding a Vicious Cycle:
Do Out-of-School Suspensions Impact Student Test Scores?**

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944346

Kaitlin P. Anderson (kaitlina@uark.edu)

Gary W. Ritter, Ph.D.

Gema Zamarro, Ph.D.

Response to NEPC (National Education Policy Center) Review of Working Paper posted at:

<http://nepc.colorado.edu/thinktank/review-discipline>

Reviewers: Brea L. Perry, Daniel J. Losen

Review Posted: June 1, 2017

Our Response Posted: June 7, 2017

Response to June 1, 2017 NEPC Review of:
**Understanding a Vicious Cycle:
Do Out-of-School Suspensions Impact Student Test Scores?**

Kaitlin P. Anderson, Gary W. Ritter, Gema Zamarro

University of Arkansas, College of Education and Health Professions

Response Submitted: June 7, 2017

We'd like to thank NEPC for reviewing our paper. This working paper, "[Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?](#)," is one of several studies we have conducted on the topic of school discipline. In our other analyses, including a [paper](#) recently published at *Educational Policy Analysis Archives*, we have used infraction-level data to document the disparate disciplinary consequences imposed upon Black students.

Presumably, NEPC chose to review this working paper because of its counterintuitive findings. While several of the criticisms levied in the review will ultimately improve our paper, we do believe there are several areas in which the NEPC review team missed the point and did not fairly represent our analysis or our findings. For the sake of clarity, we will focus on three points of departure where we find fault with the substance of the NEPC review.

First of all, the review team claims that our results run counter to "robust research concluding that exclusionary discipline has a harmful academic impact"; in our view, the research considering the causal impacts of exclusionary discipline is awfully thin and studies such as ours, which use infraction-level data, are needed.

While it is certainly true that a long line of work has found that suspensions are associated with low academic achievement, most of the work is correlational. Even the five studies (Arcia, 2006; Davis & Jordan, 1994; Marchbanks et al., 2015; Kirk & Sampson, 2013; Skiba & Rausch, 2004) cited by Perry and Losen in their review of our work - which they claim includes experimental designs - are quasi-experimental (e.g. matching designs) at best. Indeed, one motivation for our interest in this topic is the great challenge of disentangling causal relationships from endogenous correlations in the area of discipline work.

Any existing relationship between suspensions and student achievement can be explained in numerous ways. Perhaps a suspended student later chose to disengage from school or simply fell behind due to time away from class and then suffered decreased performance. Or, perhaps the student began to fall behind academically and then began to act up in class because he was disengaged academically. Or, struggles outside of school began to influence the student in school, causing him to simultaneously struggle academically and misbehave. Unfortunately, most of the existing analyses of the relationship between school discipline and student achievement do not allow the researchers to distinguish among the above interpretations.

Specifically, most prior studies do not allow the authors to control for infractions and the student's infraction history while examining the impact of disciplinary consequences on student achievement. We believe this is necessary to obtain causal effects as it allows us to compare students with similar infraction histories that potentially are also similar in other important unobservables that could drive a spurious correlation between discipline actions and test scores. Thus, to conduct analyses that properly address the endogeneity issues described above, we believe researchers need excellent data and perhaps complicated methods. Thus, while the review team questions the "face validity" of our findings because they run counter to the "weight of the evidence", we argue the existing evidence tells us little, if anything, about causal impacts of discipline and our work represents an important contribution.

Second, the NEPC review team questions our econometric methods and critiques, for example, our "lagged-year" strategy for exploring the effects of exclusionary discipline on student performance. The review team writes that "it is incongruous that the paper measures the effects of days of OSS on test scores at least a full year after the suspension occurred". Here, it seems, the reviewers are critiquing us for asking the wrong question, or asking a different question than they would ask. However, there are two reasons why we think this is exactly the right question to ask. One reason is that a longer term impact is arguably more important than a short term impact. That is, we should be very concerned with the impact of exclusionary discipline if it persisted beyond the school year in which the suspension occurred. More importantly, however, asking about the relationship in this way addresses, to the extent possible, the endogeneity concerns inherent in analyzing the relationship between student discipline and academic performance (we describe this carefully in pages 8-12 of our working paper).

The review team also criticizes the fact that we control for infraction type and infraction history in our model, even while this is one of the virtues of our data and analytic strategy. This granular level of information allows us to ask, even after accounting for whatever misbehavior was cited, what is the impact of receiving exclusionary discipline (where the counterfactual is some other, non-exclusionary consequence or no consequence). This strategy works because, for the same types of infractions, some students do receive exclusionary discipline and some do not. That the review team criticized this analytical choice suggests a misunderstanding of our strategy.

Third, and finally, the review team finds fault with our implications; unfortunately, in this case, the reviewers misrepresented our conclusions rather than simply stating our views. In section 3 of the NEPC review, the reviewers state, incorrectly, that the “*paper essentially concludes that punishment works.*” And then, after assigning to us a conclusion that we did not offer, the reviewers go on to highlight the stupidity of this conclusion! However, this is not at all what we say or conclude. Rather than incorrectly paraphrase our conclusions, the review team could have simply quoted one of our concluding lines from page 26, where we write that our results do not suggest a positive impact, but they do demonstrate that we do not find “a negative causal impact of OSS on test scores.”

In sum, the NEPC review team provides some helpful guidance and cautions us that we should be even clearer on the exact nature of the lessons learned from these results. The tone of the review is unnecessarily harsh, as the reviewers refer to our attempt at sorting out causal effects as “irresponsible science, masked by very sophisticated methods that stack the deck in favor of finding results that cast doubt on the value of discipline reform.”

Any careful reader of our working paper would see that this is untrue and we do indeed believe that there is need for discipline reform; we write in our conclusion (pp. 29-30) that:

“However, there could be other reasons that school leaders may want to use exclusionary discipline sparingly. There is much evidence that exclusionary discipline disproportionately affects students of color ... Perhaps, regardless of the positive or null impacts on student test scores, if perceived as overly harsh or unfair, the use of exclusionary discipline could still lead to negative school climate or distrust in a school community.”

Rather than “cast doubt on the value of discipline reform,” we instead argued that, if we expect academic outcomes for suspended students to improve, reducing reliance on exclusionary discipline must also be accompanied by “improving access to preventative and supportive systems at the same time.”

Given that the NEPC review team somehow overlooked this and attributed conclusions to us that we did not draw, we will clearly state here what we conclude from this analysis. Our prior belief was that discipline reform was a good idea, and we expected to find that exclusionary discipline would lead to depressed academic outcomes. While we were surprised that we did not find this relationship, we still believe that school discipline reform is a good idea for the reasons stated above. However, this careful analysis does suggest that policymakers should not expect academic gains (at least in terms of student test scores) to follow directly from decreases in the use of exclusionary discipline.