



NEPC

NATIONAL EDUCATION
POLICY CENTER

NEPC REVIEW: *UNDERSTANDING A VICIOUS CYCLE: DO OUT-OF-SCHOOL SUSPENSIONS IMPACT STUDENT TEST SCORES?*

Reviewed By:

Brea L. Perry, Indiana University
Daniel Losen, University of California Los Angeles

June 2017

Summary of Review

A new report from the Department of Education Reform at the University of Arkansas examines the association between out-of-school suspensions and student test scores. Using dynamic and multilevel regression modeling of six years of school records from all K-12 public schools in Arkansas, the paper purports to estimate a causal relationship between exclusionary discipline and academic performance. It concludes, in contrast to prior work, that the number of days of suspension a student receives has a positive relationship to math and language arts test scores. However, the outcomes (end-of-year test scores) are measured at least a full year after the hypothesized causal factor (days suspended), and there is no control for days suspended in the year the test was taken. Consequently, the results do not reflect missed instructional time for the tested material or other associated mechanisms through which suspension might adversely affect test scores. The analyses also control for a large number of infractions that are strongly related to days of out-of-school suspension, which may produce biased results. The findings also have weak face validity, in light of the weight of evidence suggesting that exclusionary discipline and school absences have adverse effects on test scores, GPA, grade retention, and dropping out. For these reasons, the reviewers caution that this paper is not useful for policymakers.



Kevin Welner
NEPC Director

William Mathis
Managing Director

Alex Molnar
Publishing Director

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

This review is one of a series made possible in part by funding from the Great Lakes Center for Education Research and Practice.



<http://www.greatlakescenter.org>
GreatLakesCenter@greatlakescenter.org

NEPC REVIEW: *UNDERSTANDING A VICIOUS CYCLE: DO OUT-OF-SCHOOL SUSPENSIONS IMPACT STUDENT TEST SCORES?*

Brea L. Perry, Indiana University
Daniel Losen, University of California Los Angeles

I. Introduction

Rates of out-of-school suspension (OSS) and expulsion have doubled since the 1970s, with nearly three million students suspended from school in the U.S. in 2010 alone.¹ The increase is attributable to a shift in disciplinary policy in the education system that borrows ideologies and techniques from the criminal justice system, including zero tolerance, random searches, security cameras, and uniformed police officers in schools. At the same time, researchers and policymakers have become increasingly concerned about the consequences of punitive educational environments for students, documenting negative effects of exclusionary discipline on student success using a variety of metrics and methods, including experimental designs.² Drawing on a wealth of empirical evidence, critics argue that suspension and other punishments that remove children from school are overused, unequally distributed, and ineffective. In response, the U.S. Department of Education (DOE) released new school discipline guidelines in 2014 directing school administrators to use exclusionary discipline only as a last resort.

The paper reviewed here – *Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?* – addresses the consequences of exclusionary discipline³, and is published as a working paper⁴ by the University of Arkansas Department of Educational Reform (EDRE). This paper begins with the claim that the body of empirical research that prompted the DOE recommendations is merely correlational. The goal of the paper is to improve upon existing work using sophisticated statistical modeling in an attempt to more closely approximate a causal test of the relationship between exclusionary discipline and academic performance. In other words, does being suspended actually cause a student to perform worse in school? The study attempts to eliminate the possibility that suspension and low achievement are correlated because the kinds of students that violate school rules are also those that are likely to perform poorly, or because students in districts with high suspension rates also have poorer performing schools. The paper measures the lagged impact of suspension on test scores, using the justification of needing to control for the possibility that both bad behavior and lower scores in the same year are caused by an exogenous shock. To determine whether suspensions in one year cause lower achievement the following year, the paper presents findings from a longitudinal and multilevel regression analysis of six years of school records from all K-12 public schools in Arkansas.

II. Findings and Conclusions of the Report

The paper addresses two distinct but related research questions. First, what is the impact of days of out-of-school suspension on academic performance as measured by standardized tests in language arts and math administered near the end of the academic year after the suspension occurred? Results obtained from traditional regression models (i.e., Ordinary Least Squares regression) are compared to those from more advanced fixed effects regression models. The latter method takes into account individual differences that might affect both a student's likelihood of being suspended and their academic performance. The traditional OLS regression methods indicate a small but significant negative association between the number of days a student was suspended and the student's scores on end-of-year math and language arts tests given the following year. Despite this lengthy gap between suspensions and test outcomes, the initial results are consistent with robust research concluding that exclusionary discipline has a harmful academic impact. However, the more advanced "preferred" models suggest a different result. Namely, increasing days of out-of-school suspension are associated with a very slight improvement in test scores at the end of the following school year. The researchers acknowledge that the relationship is negligible in magnitude, yet statistically significant.

The second question examined is whether OSS is more or less beneficial for any of the following six subgroups compared with their counterparts: students who are free/reduced lunch eligible, non-white, boys, placed in special education, low performing, and in elementary school (versus middle or high school grouped together). To compare across groups, models are run separately for each group. The working paper concludes, "The most consistent results from the subgroup analyses are that regular education students' and non-white students' test scores consistently get a boost from OSS" (p. 19), though the reported differences in the effects of suspension days across these groups are not statistically significant. In short, the data indicate that the effects of OSS do not vary across race, gender, social class, age, or special education status.

III. The Report's Rationale for Its Findings and Conclusions

The paper essentially concludes that punishment works, resulting in very small improvements in academic performance. The researchers argue that advanced statistical modeling produces more accurate estimates of the causal impact of OSS on achievement, and therefore their findings weaken the academic rationale against the use of suspension. Several possible explanations for their findings are raised, including the possibility that exclusionary punishments work as intended by encouraging students to "get back on track" (p. 20). Alternatively, the paper suggests that suspension may trigger grade retention, which the researchers imply is a positive academic outcome. This rationale contradicts a meta-analysis of studies on grade retention suggesting long-term academic harms.⁵ The paper also offers implications for policymakers, cautioning that academic gains will not follow the implementation of policies limiting use of exclusionary discipline or mandating less punitive approaches. Rather, the paper offers claims from studies that are neither causal nor peer-reviewed to suggest

that unintended negative consequences of such policies may occur – specifically, the development of a disorderly, unsafe, and drug-filled environment that is harmful to learning.⁶

IV. The Report’s Use of the Research Literature

The use of the research literature in this paper is appropriate in scope, and many relevant studies are cited. One exception is in a reliance, in the discussion of policy implications, on two reports produced by conservative think tanks that are not peer-reviewed⁷. More importantly, we find that critical methodological choices made in the paper (described in detail in Section V below) ignore insights from previous research. For example, the literature on absenteeism speaks to the timing of missed instructional days, which may be one factor driving the relationship between exclusionary discipline and academic performance. Researchers found that fourth grade students who missed three or more days in the month before a standardized test scored a full grade level below students with similar characteristics, on average.⁸ In light of this and similar studies, 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.

Similarly, the decision to focus on test scores as the sole measure of academic performance also contradicts the existing literature, though this is mentioned as a limitation. Researchers in Texas with an equally robust longitudinal data set and after controlling for over 40 variables found that being suspended predicted a 92% increase in the odds of grade retention, and a 24% increase in the odds of dropping out.⁹ Using the same Arkansas dataset, for example, researchers (including an author of the paper reviewed here) found that suspension is associated with an increase in the likelihood of grade retention.¹⁰ This result is inconsistent with the notion that out-of-school suspensions improve academic performance, since most grade retentions are a direct consequence of falling behind academically.¹¹ Illogically, however, the paper cites the results on grade retention as supporting evidence, suggesting that students scored better on tests because they had an additional year to master the material. This result is not consistent with the literature’s conclusion that suspension has a negative relationship with meaningful or long-term academic performance.

V. Review of the Report’s Methods

The major shortcoming of this study is the examination of the effects of out-of-school suspension on tests administered at the end of the following academic year, permitting the passage of as much as two years between the “treatment” (i.e., suspension) and the outcome. The rationale provided is that a life event (e.g., parental divorce) occurring in the same year as both the suspension and the test could be the cause of both the lower achievement and the problematic behavior. However, this approach overlooks the most logical impact of out-of-school suspension on achievement; namely, that a suspended student scored lower due to missed instruction in the material evaluated by the test. Some of the test scores used as

outcome variables are course-specific, assessing mastery of material presented in the current school year for a given course (e.g., geometry, algebra). Yet, this analysis examines the effect of missed instructional days in another course, taken the previous year, on the current year's test performance. Moreover, all of the more immediate effects of being suspended that might cause lower test scores – such as stigma and stereotype threat, increased stress, reduced self-esteem, negative relationships with teachers, resentment toward administrators, perceptions of discrimination, etc. – are not captured or are likely to be weakly captured.

Put differently, the one-year lag between the suspension and the test may miss the “treatment” effect altogether and instead capture the start of a recovery from a suspension that occurred in the prior year. Logically, we would expect the largest impact of suspension to be in the year it occurred, with students doing better academically in years they were not suspended, including the year after a suspension. However, in this analysis, students that were not suspended in year one, but were suspended in the tested year (year two), were regarded as “non-suspended” for the purpose of estimating the impact of suspension on that student's performance in year two.

In other words, this analysis assumes not only that we should see stronger effects of suspension the year after it occurs, but it does not consider any impact that suspension might have at the end of the same year.

Another problem is that students who were expelled or placed in alternative schools, at some point, are omitted from the analysis. The paper argues that this decision provides an estimate of the effects of suspension for more typical students rather than for “extremely misbehaving” students (p. 8). To the extent that students who struggle academically are more likely to also have behavioral challenges, this control minimizes the achievement impact. Arguably, students with greater needs are more strongly affected by being excluded from school than a typical student, who might more easily recover from missed instructional time and the stigma of suspension. This criticism – that the omitted students are more likely to suffer adverse consequences of exclusionary discipline – is supported by the results of a robustness test performed to assess the impact of dropping these students from the analysis. Specifically, the tables in the paper indicate that the effects of OSS are null or negative, not positive and significant, when these more harshly punished students are taken into account.

In short, while the researchers expressly state that their study comes closer to providing causal evidence than previous research, the decisions to use a one-year lag without a control for current year effects of suspension and also to exclude “extremely misbehaving” students (p. 8) defy common sense. Together these problems undermine the validity of the findings. This research design is analogous to a drug company examining the potential negative side effects of a drug by only measuring the effects that lingered for more than a year, and by excluding from their analysis all the patients that died or suffered very serious harm the same year they took the medicine being evaluated.

An additional point of concern is the inclusion of a large number of fixed effects and variables that are very closely associated with suspension. In particular, the models in the paper include controls for a series of 12 different variables representing distinct types of infractions. The length of an out-of-school suspension in Arkansas¹², and in most states, is directly

related to the specific offense. By law, school administrators must adhere to minimum and maximum penalties, and some offenses are grounds for mandatory suspension of a minimum length. As a result, controlling for a count of each specific type of infraction is tantamount to controlling for one or more out-of-school suspensions of a specific average length. For example, if a theft results in an average of three days of suspension, controlling for the number of thefts a student commits has the de facto effect of controlling for the number of three-day suspensions. Any observed effect of suspension above and beyond the number of specific offenses reflects an unusually small or large number of OSS days for that offense, not the effect of the suspension experience, per se. This may be creating a problem called overfitting, which occurs when the model is much more complicated than it needs to be to explain what is happening in the data. When regression models are overfitted, as we suspect these are, they do an excellent job of describing the existing data, but will do a terrible job of predicting new data. Clearly, poor predictions would not be a sound foundation for adopting or changing policies.

We find the procedures used to conduct and report the statistical analyses for this paper to be inadequate.

Along the same lines, the R-squared values for the advanced models published in the original version of the working paper suggest that the results may not be trustworthy. The R-squared values were reported in the version of the paper uploaded on March 30th, but removed from a revised version uploaded on April 29th without explanation. The R-squared is an indication of the proportion of variation (here, differences in test scores within students over time) that is explained by the model. For example, an R-squared of 0.35 indicates that 35% of change in the dependent variable is explained by changes in the independent variables, but 65% is left unexplained. The R-squared value for the preferred models reported in the early version of the paper is 0.000 for virtually every model. This signals that despite the appearance of a statistically significant “effect,” the models do not explain any change in test scores over time.

More generally, we find the procedures used to conduct and report the statistical analyses for this paper to be inadequate. Particularly in cases where results are inconsistent with previous research, it is critical to triangulate conclusions using multiple models and to conduct a range of tests to ensure that the findings are not simply an artifact of particular analytic decisions or idiosyncrasies in the data. The paper does not do due diligence in this regard. Along the same lines, there is insufficient information provided in the paper to reasonably evaluate important features of these data and statistical models. For example, basic descriptive statistics for key variables, including OSS days and other discipline outcomes and infractions, are not reported. Also, results pertaining to the effects of all control variables are, for some unknown reason, omitted from tables and text. This information is required to assess the credibility of model results. Finally, results that are inconsistent with a null or positive effect of OSS days on test scores are dismissed as “reversion to the mean” (p. 17) or are presumed to be statistically significant by “chance occurrence” (p. 20). Moreover, careful statements in some sections of the paper (e.g., choosing to describe the causal impact as negligible, rather than positive) are offset by exaggerated and inaccurate reporting of findings in others (e.g., reporting that non-white students get a boost from out-of-school suspension). These patterns are inconsistent with professional norms governing the responsible reporting of research.

VI. Review of the Validity of the Findings and Conclusions

Taken together, the findings and related conclusions presented in the paper lack validity on multiple grounds. The findings are at odds with a rather substantial body of empirical research suggesting that out-of-school suspensions, and missed instructional days more generally, have negative consequences for students¹³. Additionally, this paper is characterized by a series of questionable methodological decisions (e.g., the one-year lag, omission of “misbehaving” students, and inclusion of indicators for all infractions), and explains nothing about why students who have been suspended perform better or worse on standardized tests from year to year. Finally, the researchers fail to provide access to basic descriptive and multivariate results, and at times offer conclusions that are inconsistent with their own statistical findings. In our opinion, this paper is an example of 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.

VII. Usefulness of the Report for Guidance of Policy and Practice

This paper offers no value for guiding disciplinary policy and practice. The major policy implication offered in the paper is that, given the potential negative consequences of discipline reform, we should consider scaling back these efforts. This is a straw man argument. Disciplinary reform efforts are never aimed solely at improving test scores without regard for other academic impacts. In addition to the goals of reducing grade retention, dropping out of school, and juvenile justice system involvement, there are typically broader aims of discipline reform related to achieving equality and justice for all students. In contrast, the narrow focus in this paper on only the delayed effects of out-of-school suspension, while ignoring more plausible and proximal mechanisms, puts any stated policy implications on shaky ground.

Notes and Resources

- 1 Losen, D.J. (2011, October). *Discipline policies, successful schools, and racial justice*. Boulder, CO: National Education Policy Center.
- 2 Arcia, E. (2006, May). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society*, 38(3), 359-369.

Davis, J.E., & Jordan, W.J. (1994, Autumn). The effects of school context, structure, and experiences on African American males in middle and high schools. *Journal of Negro Education*, 63(4), 570-587.

Marchbanks III, M.P., Blake, J.J., Smith, D., Seibert, A.L., Carmichael, D., Booth, E.A., & Fabelo, T. (2015, February). More than a drop in the bucket: The social and economic costs of dropouts and grade retentions associated with exclusionary discipline. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 5(2), 17.

Kirk, D.S., & Sampson, R.J. (2013, January). Juvenile arrest and collateral educational damage in the transition to adulthood." *Sociology of Education*, 86(1), 36-62.

Skiba, R.J. & Rausch, M.K. (2004, July). *The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP scores*. Bloomington, IN: Center for Evaluation and Education Policy.
- 3 Kaitlin P. Anderson, Gary W. Ritter, Gema Zamarro (2017) *Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?* EDRE Working Paper 2017-09. Retrieved April 25, 2017, from: <http://www.uaedreform.org/understanding-a-vicious-cycle-do-out-of-school-suspensions-impact-student-test-scores/>
- 4 A difference between a working paper and a journal article or report is that the former can be updated periodically online using the same citation. This particular paper was first made public on March 31st, 2017 and updated on April 30th, 2017. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944346##
- 5 Jimerson, S. (2001, September). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420-438.

McCoy, A.R., & Reynolds, A.J. (1999, Autumn). Grade retention and school performance: An extended investigation. *Journal of School Psychology*, 37(3), 273-298.
- 6 Loveless, T. (2017, March). *The 2017 Brown Center report on American education. How well are American students learning?* Washington, DC: Brookings Institution.

Eden, M. (2017, March). *School discipline reform and disorder: Evidence from New York City Public Schools, 2012-16*. New York, NY: Manhattan Institute.
- 7 Loveless, T. (2017, March). *The 2017 Brown Center report on American education. How well are American students learning?* Washington, DC: Brookings Institution.

Eden, M. (2017, March). *School discipline reform and disorder: Evidence from New York City Public Schools, 2012-16*. New York, NY: Manhattan Institute.
- 8 Ginsburg, A., Jordan, P., & Chang, H. (2014, August). Absences add up: How school attendance influences student success. *Attendance Works*. Retrieved May 5, 2017, from http://www.attendanceworks.org/wordpress/wp-content/uploads/2014/09/Absences-Add-Up_September-3rd-2014.pdf

- 9 Marchbanks III, M.P., Blake, J.J., Smith, D., Seibert, A.L., Carmichael, D., Booth, E.A., & Fabelo, T. (2015, February). More than a drop in the bucket: The social and economic costs of dropouts and grade retentions associated with exclusionary discipline. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 5(2), 17.
 - 10 Swanson, E., Erickson, H.H., & Ritter, G.W. (2017, April). *Examining the impacts of middle school disciplinary policies on 9th grade retention*. EDRE Working Paper 2017-11. Retrieved May 5, 2017, from <http://www.uaedreform.org/examining-the-impacts-of-middle-school-disciplinary-policies-on-9th-grade-retention/>
 - 11 Beebe-Frankenberger, M., Bocian, K.M., MacMillan, D.L., & Gresham, F.M. (2004, June). Sorting second-grade students: Differentiating those retained from those promoted. *Journal of Educational Psychology*, 96(2), 204-215.
 - 12 Department of Education. (2016, January). *Arkansas Compilation of School Discipline Laws and Regulations*. Retrieved May 10, 2017, from <https://safesupportivelearning.ed.gov/sites/default/files/discipline-compendium/Arkansas%20School%20Discipline%20Laws%20and%20Regulations.pdf>
 - 13 Losen, D.J. (2011, October). *Discipline policies, successful schools, and racial justice*. Boulder, CO: National Education Policy Center.
- Arcia, E. (2006, May). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society*, 38(3), 359-369.
- Davis, J.E., & Jordan, W.J. (1994, Autumn). The effects of school context, structure, and experiences on African American males in middle and high schools. *Journal of Negro Education*, 63(4), 570-587.
- Marchbanks III, M.P., Blake, J.J., Smith, D., Seibert, A.L., Carmichael, D., Booth, E.A., & Fabelo, T. (2015, February). More than a drop in the bucket: The social and economic costs of dropouts and grade retentions associated with exclusionary discipline. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 5(2), 1.
- Balfanz, R., Byrnes, V., & Fox, J. (2015). Sent Home and Put Off Track: The Antecedents Disproportionalities, and Consequences of Being Suspended in the 9th Grade. In D. Losen (Ed.), *Closing the School Discipline Gap*. New York, NY: Teachers College Press.
- Kirk, D.S., & Sampson, R.J. (2013, January). Juvenile arrest and collateral educational damage in the transition to adulthood." *Sociology of Education*, 86(1), 36-62.
- Skiba, R.J. & Rausch, M.K. (2004, July). *The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP scores*. Bloomington, IN: Center for Evaluation and Education Policy.

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

AUTHORS: Kaitlin P. Anderson, Gary W. Ritter, & Gema Zamarro

PUBLISHER/THINK TANK: The University of Arkansas, Department of Education Reform

DOCUMENT RELEASE DATE: March 30, 2017

REVIEW DATE: June 1, 2017

REVIEWERS: Brea L. Perry, Indiana University
Daniel Losen, University of California Los Angeles

E-MAIL ADDRESSES: blperry@indiana.edu; losendan@gmail.com

PHONE NUMBERS: (812) 856-0447; (781) 861-1222

SUGGESTED CITATION:

Perry, B.L. & Losen, D. J., (2017). *Review of "Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?"* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-discipline>