Zero Tolerance and Zero Statistical Significance: ZTPs and High School Graduation Rates

Introduction

‘Zero tolerance’ policies are school policies that have pre-mandated and harsh consequences for student misconduct. Beginning in the 1980s as a way to curb violence and drug usage in schools, the term ‘zero tolerance’ has become an umbrella term for the rigid and punitive measures used to discipline students all across the United States (Curan, 2019). A recent study by the Brookings Institution revealed that about 62% of public schools in the nation used zero tolerance policies to facilitate disciplinary action during the 2021-22 school year (Perera & Diliberti, 2023).

The widespread use of zero tolerance policies suggests that they may be effective in controlling student behavior and improving their overall educational experience. However, as with any disciplinary policy, this claim warrants further investigation. Specifically, with other disciplinary measures available, the question of how successful zero tolerance policies are in helping students graduate is uniquely important. In other words, what exactly is the causal effect of zero tolerance policies on high school graduation rates?

Countless studies have noted the correlation between educational attainment and income levels, as having a ‘higher’ degree allows one to access higher pay scales (Tamborini et al.,
2015) (Taylor et al., 2019). As having completed high school education (or some equivalent) allows one access to higher education (in the form of a bachelor’s degree or similar), it seems a reasonable benchmark for considering outcomes. Supporters of zero tolerance policies argue that zero tolerance policies (ZTPs for short) raise the stakes on bad behavior, creating an environment in which students are less likely to act out so that more value can be gained from their time in school. While this argument makes intuitive sense, those who oppose ZTPs make the case that such policies keep students out of school through the overuse of suspensions and have potential to be exploited against students of minoritized identities (American Psychological Association, 2008). They advocate for alternative measures that do not default to suspension, expulsion, or other disciplinary equivalents.

For our research, we begin with the hypothesis that zero tolerance policies lower student outcomes, making them less likely to graduate and lowering the overall graduation rate across the state. In this context, being ‘without’ a zero tolerance policy simply means requiring some alternative to the default, predetermined consequence outlined for student misconduct.

Rather than simply comparing the outcomes of those with zero tolerance policies to those without, a difference-in-differences approach offers a more nuanced analysis of the issue. The present study uses panel data to compare the change in state-wide graduation rates for those who adopted zero tolerance policies against those who did not. By assuming all states would generally trend the same way had no changes been made (parallel trends assumption), this approach attempts to isolate the effects of zero tolerance policies on graduation rates to the extent possible.

Ultimately, our findings are statistically insignificant and therefore inconclusive. We find no significant changes to high school graduation rates at the state level when comparing those
states without zero tolerance policies to those without. There are many reasons why this may be the case in our analysis, and as the area of study is generally lacking in empirical evidence, further research may be beneficial.

**Literature Review**

Prior research on zero tolerance policies has been largely qualitative and has primarily focused on the impact on expulsion rates and school safety. Generally, this research has found that zero tolerance policies worsen student outcomes, with impacts unequally distributed on the basis of race. The bulk of this research was conducted in the early 2000s, following the initial implementation of zero tolerance policies in the 1980s and 1990s.

A report from The Civil Rights Project (2000) provides an overview of the association between the implementation of zero tolerance policies and trends in suspension rates. Citing state-reported data, the report notes that Wisconsin experienced a 34% increase in suspensions between the years 1992 and 2000, and that Chicago Public Schools saw an increase from 14 expulsions per year during the 1991-1992 school year to 737 in 1998-99 (p.3). Skiba and Peterson (1999) provide more context for these numbers. Compiling reports from national media, Skiba and Peterson give a sampling of the kinds of incidents that, under zero tolerance policies, have resulted in suspensions. Many of these infractions were relatively minor and posed no clear risk to school safety or climate (e.g., flashing a toy gun in class, sharing zinc cough drops, or sharing an inhaler) (p.375).

Given the relative triviality of many punishable transgressions under zero tolerance policies, other literature in the field has sought to answer two questions: 1) whether racial bias plays a sizable role in suspensions and 2) whether or not these suspensions actually increased
school safety. Zero tolerance policies are predicated on the idea that the threat of suspension deters students from acting in disruptive and dangerous ways. However, a study by Huang and Cornell (2021) found that students are often targeted in a biased way and that student-reported measures of school safety are negatively correlated with teacher support for zero tolerance policies. Running a linear regression on data from the 2016 Virginia Secondary School Climate Survey (which included responses from 110,889 students) Huang and Cornell determined that teacher support for zero tolerance policies is, on average, associated with a .22 percentage point increase in expulsion for black students ($p < .05$) as opposed to only a .10 percentage point increase for white students ($p < .05$) (p.397). Furthermore, a 10 percentage point increase in support for zero tolerance was associated with a 4 percentage point reduction in student ratings of school safety (p. 399).

In light of these observed increases in expulsion rates, apparent racial discrimination, and suggested decreases in student-reported measures of safety, it seems important to consider the long-term implications of zero tolerance policies. It is for this reason that we have chosen to study the impact of zero tolerance policies on four-year graduation rates. We view this as a reasonable outcome variable because empirical analysis draws a clear connection between suspension rates and high school graduation rates. This link was established in a research paper by Losen and Rumberger (2017) who ran a multivariate regression (controlling for grades and tests scores) on a dummy variable for whether an individual had been suspended. They found that in California, students who had been suspended had, on average, a graduation rate that was 6.5 percentage points lower than those who had not been suspended (p.4).

High school graduation rates are a useful outcome variable because they have significant long-term impacts, which is what we seek to measure. Harvard economists Richard Murnane
notes that high school graduates consistently see higher wages than non-high school graduates, at about 75 cents per dollar (Murnane, 2013, 31). In addition, Baher-Hicks et al. (2019) capitalized on changes in school assignment to regress adult crime rates on a measure of the strictness of disciplinary practices, which are correlated with lower graduation rates. Utilizing neighborhood and cohort fixed effects (p.15), these researchers found that students who moved to these stricter schools were 2.5 percentage points more likely to have been incarcerated (p < 0.05) than those who were not reassigned to these schools (p.18).

Data

The two primary variables of interest for our study are the existence of a policy requiring alternatives to suspension and high school graduation rate, both at the state-year level.

For the purposes of this paper, we use the existence of a state-wide policy mandating the use of alternatives to suspension as an indicator of whether the state is implementing zero tolerance policies in its schools. Data concerning the existence of state-wide policies requiring the use of alternatives to suspension was obtained from Law Atlas, a policy surveillance website affiliated with Temple University’s Beasley School of Law. Specifically, the Law Atlas Policy Surveillance Program offers mapping and data related to state policy decisions, including school discipline policies. The Law Atlas school discipline database contains data about the existence of policies requiring use of alternatives to suspension from 2008 to 2018. Observations exist at the state-year level for forty-eight states (excluding North Dakota and Ohio) and the District of Columbia, for a total of forty-nine included localities. Due to constraints in the high school graduation rate data, only observations from 2010 to 2018 were included. Each state is categorized as either requiring the use of alternatives to suspensions, recommending the use of
alternatives to suspension, or neither requiring nor recommending the use of alternatives to suspension. Because this paper is concerned with the causal effect of requiring the use of alternatives to suspension, we categorize states as either requiring the use of alternatives or not requiring their use, with states that recommend but do not require alternatives and states that neither recommend nor require the use of alternatives in the same category.\(^1\) The variable alt_required is an indicator for whether states require the use of alternatives to suspension, with a value of 1 indicating that they do require the use of alternatives, while a value of 0 indicates that they do not require the use of alternatives. The mean value of alt_required is .23, indicating that 23\% of the state-year observations show states requiring the use of alternatives to suspension (Table 1). Additionally, we use the variable alt_expansive, which has a value of 1 if a state required or recommended the use of alternatives to suspension in a given year and a value of 0 if a state neither requires or recommends the use of alternatives to suspension in a given year. Alt_expansive has a higher mean of .37, indicating that 37\% of state-year observations have states requiring or recommending the use of alternatives to suspension. Since alt_expansive is a broader category than alt_required, the higher mean is not surprising.

The second major variable identified for this paper is high school graduation rate, also at the state-year level. Data was obtained from the National Center for Education Statistics, which collects data for the Department of Education. The measure of high school graduation rate used for this paper was the public high school four-year adjusted cohort graduation rate (ACGR), which tracks graduation rate within four years of beginning high school, while adjusting for students transferring in or out of a state. The ACGR is available for all fifty states and the

\(^1\) One state proved particularly challenging to categorize. Arkansas originally required the use of alternatives to suspension, but they changed their law to no longer require the use of alternatives to suspension in April of 2011. By July of 2011, they returned to requiring the use of alternatives, a policy that remained in place until 2012. Since alternatives were required for the majority of 2011, we categorized Arkansas in 2011 as requiring the use of alternatives to suspension.
District of Columbia from 2010 to 2019, but only data from 2010 to 2018 was used due to limitations in the data about state policies requiring use of alternatives to suspension. While the ACGR is not available for each state in each year, data exists for the overwhelming majority of state-year combinations (453 of the 459 state-year combinations). ACGR is measured in percentage points, on a scale of 0 to 100. Across the observed state-year combinations, average ACGR was 82.63, representing 82.63% of students in ninth grade graduating high school within four years. Across the observations, the minimum observed value was a graduation rate of 59%, and the maximum observed value was 92% (Table 1).

**Identification Strategy**

We use the difference-in-differences identification strategy to exploit changes in state policy on requiring the use of alternatives to suspension throughout the period of 2010-2018. Over the course of our available data, four states changed their policies. This is a relatively small number, which limits the effectiveness of the analysis. By comparing trends in states that changed their policies and states that did not change their policies over time, we hope to find the causal effect of requiring the use of alternatives to suspension while using state and time fixed effects. This allows us to take into account inherent differences in graduation rate by state and year to find the causal effect of a state changing its policy.

The major assumption for difference-in-differences is the parallel trends assumption. We assume that states that do not change their policy are reasonable counterfactuals for states that do change their policies. This assumption is reasonable, if imperfect. There are certainly potential events that could affect graduation rates in one state, but not in all states. For example, if a natural disaster were to strike one state that did not change its suspension policy and cause a
much lower graduation rate that year, the trend of decreasing graduation rates would not be an accurate counterfactual for a state that did change its suspension policy.

However, one benefit to the comparatively small number of states that changed their policies is that there are a huge number of states that together serve as the counterfactual for states that did change their policies. It is unlikely that all of the states that did not change their policies experienced some kind of simultaneous freak event that shifted graduation rates, while the states that did change their policies did not experience the same freak event. Rather, if some unusual event affected graduation rates for a significant number of states, it was probably a national event that also affected states that changed their policies (e.g. global pandemic). We are confident enough that the parallel trends assumption holds to progress with difference-in-differences analysis.

**Regression Specification**

Because we are using the differences-in-differences identification strategy, our regression of state graduation rates on alt_required takes the following form:

\[
\text{grad_rate}_{st} = \alpha + \beta_{DD} \text{alt_required}_{st} + \beta_{AK} A K_s + \ldots + \beta_{WY} W Y_s + \beta_{2010} Year2010_t + \ldots + \beta_{2018} Year2018_t + \epsilon_{st}
\]

\( \text{grad_rate}_{st} \) represents the expected graduation rate in a given state during a given year. \( \text{AK}_s + \ldots + \text{WY}_s \) are a set of binary variables identifying the locality each observation belongs to, while \( \text{Year2010}_t + \ldots + \beta_{2018} Year2018_t \) are a set of binary variables identifying the year
each observation was taken in. For both the state binary variables and the year binary variables, one state/year is excluded to avoid perfect multicollinearity. This regression can also be written in the following form:

\[
\text{grad}_{st} = \alpha + \beta_{DD} \text{alt}_\text{required}_{st} + \gamma_s + \delta_t + \epsilon_{st}
\]

In the above form, \(\gamma_s\) is a vector of state fixed effects, while \(\delta_t\) is a vector of year fixed effects. Our main coefficient of interest is \(\beta_{DD}\), which represents the causal effect of requiring the use of alternatives to suspension on graduation rate. We expect \(\beta_{DD}\) to be positive, as prior literature indicates that harsh discipline policies decrease student achievement. We estimate the regression using OLS and robust standard errors. Our regressions of state graduation rate on alt_expansive take the same form, with alt_expansive substituted in as the main explanatory variable instead of alt_required.

**Empirical Results**

Table 2 details the results from our regression analysis, showing the coefficients, standard errors for the coefficients, number of observations, and R-squared for both alt_required (alternatives to suspension required) and alt_expansive (alternatives to suspension required or recommended).

The results from the regression of grad_rate on alt_required yield a coefficient of -.5020508. In the context of our model, this means that requiring alternatives to suspension lowers 4-year high school graduation rates by approximately .502 percentage points. This
contradicts the hypothesis we formed based on our literature review, which suggested that harsh school discipline laws lower student outcomes. However, it is important to keep in mind that an associated decrease of .502 percentage points is small compared to the average graduation rate of 82.63% across our data set. This means that our p-value (0.389) is very high and our results are not statistically significant at any of the commonly chosen levels of significance.

In addition to regressing grad_rate on alt_required, we also regressed grad_rate on alt_expansive. We did this to increase the number of states in our data set that changed their policy; as of 2018 (the last year in our data set) only four states had changed from not requiring alternatives to suspension to requiring alternatives to suspension, but a total of nine states had changed their policy from neither requiring nor recommending alternatives to suspension to either requiring or recommending alternatives. When we ran our regression with this new explanatory variable, two values changed significantly: 1) our coefficient became less negative, increasing from -.50205088 to -.2279079 and 2) our p-value increased from 0.389 to 0.719. This means that the size of the effect decreased and our results became even less statistically significant. One possible explanation for this is that many of the schools in states where alternatives to suspensions are “recommended but not required” did not actually alter disciplinary practices, meaning that their inclusion in our regression diluted the real effect of providing alternatives to suspensions.

The primary weakness of our model is the low variation in our explanatory variable (the small number of states that changed their policies between 2010 and 2018), which limits our ability to obtain statistically significant results. As noted, we tried to address this by expanding our explanatory variable to include states that recommend but do not require the use of alternatives to suspension. In doing this, however, we included states that may not have changed
suspension policies in practice, decreasing the size of the effect and increasing the p-value. Looking forward, one way to improve this might be to increase our number of observations (currently 435) by incorporating changes to zero tolerance policies that occur after 2018. If enough states change their policies in this additional time frame, then this increased variation in the explanatory variable might shrink our standard error and our p-value as a result.

Our model is strengthened by our use of a difference-in-differences model with state and time fixed effects. This form of regression yields more accurate results than a simple cross-sectional comparison across states, which does not control for fundamental differences between states or general time trends. The benefits of the incorporation of state and time fixed effects are reflected in our very strong R-squared, which takes a value of 0.9008 for the regression of grad_rate on alt_required and 0.9007 for the regression of grad_rate on alt_expansive. This means that for both regressions, our model explains just over 90% of the resulting graduation rate. Since R-squared is a measure of how well our regression model explains our results, this indicates that knowing state and year alone allows us to very effectively gauge outcomes to graduation rate.

**Conclusion**

This paper uses difference-in-differences analysis with state and time fixed effects to estimate the causal effect of requiring the use of alternatives to suspension on high school graduation rates in the United States. We are unable to find a statistically significant causal effect. Additionally, we are unable to find a statistically significant causal effect of requiring or recommending the use of alternatives to suspension on high school graduation rate. We are therefore unable to recommend any policy changes based on the results obtained in this paper.
As previously discussed, ZTPs remain an under-researched area of education policy. Further research could follow two potential paths. First, since 2018 (the end of the Law Atlas data), states have faced growing movements to end the school-to-prison pipeline by eliminating ZTPs. As a result, a broader data set encompassing state-wide policies until 2023 would include more states changing their policies, driving down the p-value and potentially yielding statistically significant results. The second path to further research might be using regression discontinuity design to evaluate the effect of implementing or discarding a policy requiring the use of alternatives to suspension. If researchers could find a state or locality with similar graduating cohorts before and after the change in policy, regression discontinuity design could allow them to see the effects of requiring the use of alternatives to suspension in that particular region, yielding more significant data. Regardless, more empirical research is needed in this field.
Figures and Tables

Table 1: Summary Statistics for Important Variables

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt_required</td>
<td>441</td>
<td>.238</td>
<td>.426</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>alt_expansive</td>
<td>441</td>
<td>.376</td>
<td>.485</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ACGR</td>
<td>453</td>
<td>82.63</td>
<td>6.01</td>
<td>59</td>
<td>92</td>
</tr>
</tbody>
</table>

Table 2: Regression Results

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>P-Value for Coefficient</th>
<th>N</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt_required</td>
<td>-.50205088 (.5821683)</td>
<td>0.389</td>
<td>435</td>
<td>0.9008</td>
</tr>
<tr>
<td>alt_expansive</td>
<td>-.2279079 (.6327119)</td>
<td>0.719</td>
<td>435</td>
<td>0.9007</td>
</tr>
</tbody>
</table>
References


